



NOAA Technical Memorandum NMFS-XXX-##

# GAP Production Data Documentation

Bering Sea Survey Team, Gulf of Alaska and Aleutian Island Survey Team

U.S. DEPARTMENT OF COMMERCE

National Oceanic and Atmospheric Administration  
National Marine Fisheries Service  
Northwest Fisheries Science Center



**NOAA  
FISHERIES**

# GAP Production Data Documentation

Bering Sea Survey Team<sup>1,\*</sup> and Gulf of Alaska and Aleutian Island Survey Team<sup>1,\*</sup>

1. NOAA Fisheries Alaska Fisheries Science Center, Groundfish Assessment Program

\* Correspondence: Bering Sea Survey Team afsc.gap.metadata@noaa.gov

\* Correspondence: Gulf of Alaska and Aleutian Island Survey Team afsc.gap.metadata@noaa.gov

# Table of contents

<b>I. Welcome</b>	<b>1</b>
Our Objective . . . . .	2
User Resources . . . . .	2
Background of the gap_products repo . . . . .	2
Major Advantages . . . . .	4
<b>1. Survey Background</b>	<b>6</b>
1.1. What is the research objective? . . . . .	6
1.2. Who is conducting the research? . . . . .	6
1.3. Bottom trawl surveys and regions . . . . .	7
<b>2. Workflow</b>	<b>12</b>
2.1. Data levels . . . . .	12
<b>3. News</b>	<b>14</b>
3.1. Future plans . . . . .	14
3.2. Previous updates . . . . .	14
<b>II. GAP Production Data</b>	<b>16</b>
Data Description . . . . .	17
Cite this data . . . . .	17
Data Creation . . . . .	18
<b>4. Data description</b>	<b>19</b>
4.1. Data usage . . . . .	19
4.2. Data tables . . . . .	19
<b>5. gapindex R package</b>	<b>36</b>

*Table of contents*

<b>III. AKFIN</b>	<b>37</b>
The Alaska Fisheries Information Network . . . . .	38
Cite this data . . . . .	38
<b>6. Data description</b>	<b>39</b>
6.1. Data Description . . . . .	39
6.2. Data Tables . . . . .	39
6.3. Data usage . . . . .	39
6.4. Data tables . . . . .	39
<b>7. Accessing Data</b>	<b>74</b>
7.1. Access data via Oracle (AFSC only) . . . . .	74
7.2. Data SQL Query Examples: . . . . .	74
<b>8. Access API data using R</b>	<b>91</b>
8.1. Ex. 1: Load lingcod data . . . . .	91
<b>IV. Public Data (FOSS)</b>	<b>93</b>
<b>V. Collaborators and data users</b>	<b>95</b>
Cite this data . . . . .	96
<b>9. Data description</b>	<b>97</b>
9.1. Data tables . . . . .	98
9.2. Data usage . . . . .	98
9.3. Data tables . . . . .	98
<b>10. Using the API</b>	<b>117</b>
10.1. Select and filter . . . . .	117
10.2. Select data format . . . . .	119
10.3. Run report . . . . .	121
<b>11. Access API data using R</b>	<b>122</b>
11.1. Ex. 1: Load the first 25 rows (default) of data . . . . .	122
11.2. Ex. 2: Load the first 10000 rows of data . . . . .	123
11.3. Ex. 3: Filter by Year . . . . .	123
11.4. Ex. 4: Filter by species name . . . . .	124
11.5. Ex. 5: Combination of year and name filters . . . . .	125
11.6. Ex. 6: Combination of year, srvy, stratum . . . . .	126
11.7. Ex. 7: Visualize CPUE data in distribution map . . . . .	127

*Table of contents*

<b>12. Access API data using Python</b>	<b>130</b>
<b>13. Access data using R (AFSC only)</b>	<b>136</b>
<b>14. Other products</b>	<b>139</b>
14.1. Annual Publications . . . . .	139
14.2. NOAA Fisheries Distribution Mapping and Analysis Portal (DisMAP) . .	140
14.3. R Packages . . . . .	140
14.4. Outreach . . . . .	141
14.5. Collaborators . . . . .	141
<b>VI. Contact us</b>	<b>142</b>
Suggestions and comments . . . . .	143
<b>15. Production Run Notes</b>	<b>144</b>
<b>16. R Version Metadata</b>	<b>145</b>
<b>17. Data constraints</b>	<b>147</b>
17.1. Cite this data . . . . .	147
<b>18. Access Constraints</b>	<b>148</b>
<b>19. Acknowledgments</b>	<b>149</b>
<b>20. Community Acknowledgments</b>	<b>150</b>
<b>21. Technical Acknowledgments</b>	<b>151</b>
21.1. Partners . . . . .	151
<b>22. References</b>	<b>152</b>

# List of Figures

1. Sorting and weighing fish on deck on the 2022 Bering Sea groundfish survey aboard the F/V Alaska Knight. Credit: Emily Markowitz/NOAA Fisheries. . . . .	3
1.1. Strata used in the Eastern Bering Sea Survey. . . . .	10
1.2. Strata used in the Bering Sea Slope Survey. . . . .	11
7.1. Ex. 1: GOA Pacific Ocean perch biomass and abundance. . . . .	78
7.2. Ex. 2: AI Rock sole size compositions and ridge plot. . . . .	80
7.3. Ex. 3: EBS Walleye Pollock Age Compositions and Age Pyramid. . . . .	82
7.4. Ex. 4: NBS Pacific cod biomass and abundance. . . . .	85
7.5. Ex. 5: GOA Pacific Ocean perch biomass and line plot. . . . .	88
7.6. Ex. 6: EBS Pacific Ocean perch CPUE and akgfmaps map. . . . .	90
10.1. AFSC Groundfish and Crab Assessment Program Bottom Trawl Survey data interface on the Fisheries One Stop Shop platform. . . . .	118
10.2. Diagram of selection and search tools available on the FOSS platofrom. . . . .	119
10.3. Diagram of selection and search tools available on the FOSS platofrom. . . . .	120
10.4. Diagram of the pre-set data format options. . . . .	120
10.5. Example data returned from running the report. . . . .	121
11.1. Ex. 7: Visualize CPUE data in distribution map. . . . .	129

# List of Tables

1.1. Survey summary stats . . . . .	7
7.1. Ex. 1: GOA Pacific Ocean perch biomass and abundance. . . . .	76
7.2. Ex. 2: AI Rock sole size compositions and ridge plot. . . . .	79
7.3. Ex. 3: EBS Walleye Pollock Age Compositions and Age Pyramid. . . . .	81
7.4. Ex. 4: NBS Pacific cod biomass and abundance. . . . .	84
7.5. Ex. 5: GOA Pacific Ocean perch biomass and line plot. . . . .	86
7.6. Ex. 6: EBS Pacific Ocean perch CPUE and akgfmaps map. . . . .	89
11.1.Ex. 1: Load the first 25 rows (default) of data. . . . .	123
11.2.Ex. 3: Filter by Year. . . . .	124
11.3.Ex. 4: Filter by species name. . . . .	125
11.4.Ex. 5: Combination of year and name filters. . . . .	126
11.5.Ex. 6: Combination of year, srvy, stratum. . . . .	126
11.6.Ex. 7: Visualize CPUE data in distribution map. . . . .	127

**Part I.**

# **Welcome**

### *Our Objective*

Please consider this resource to be a **Living Document**. The code in this repository is regularly being updated and improved. Please refer to releases for finalized products and project milestones.

At this time, these master production and AKFIN tables are **provisional** and we are welcoming feedback before the 2024 survey season. We look forward to hearing from you. Do not hesitate to reach out (to us at either afsc.gap.metadata@noaa.gov or GitHub issues), especially if you find discrepancies in the data or want to suggest improvements to infrastructure. Thank you in advance for your collaboration and partnership with us as we develop our future data universe.

## **Our Objective**

As part of our commitment to open science and transparency, we provide this interactive metadata guide to compliment our public-domain data. Please refer to our Draft Data Changes Brief. Once finalized, this language will be included here.

## **User Resources**

- Groundfish Assessment Program Bottom Trawl Surveys
- AFSC's Resource Assessment and Conservation Engineering Division
- Survey code books
- Publications and Data Reports
- Research Surveys conducted at AFSC

## **Background of the gap\_products repo**

This work is the result of the massive efforts of three concurrent GAP working groups:

- (1) **Index Computation Working Group:** consolidation of the methods used to produced design-based estimates of abundance and size/age composition between the Bering Sea and AIGOA survey regions.

*Background of the gap\_products repo*



Figure 1.: Sorting and weighing fish on deck on the 2022 Bering Sea groundfish survey aboard the F/V Alaska Knight. Credit: Emily Markowitz/NOAA Fisheries.

## *Major Advantages*

- (2) **Data Processes Working Group:** reorganization of the Oracle data infrastructure that houses the standard data products produced by GAP.
- (3) **Gulf of Alaska Survey Restratiification Working Group:** implementation of a new stratified random survey design in the Gulf of Alaska bottom trawl survey.

We began this effort in collaboration with the Status of Stocks team (SSMA) to present both the orientation and opportunity to interact with Gulf of Alaska data from the restratiified survey design that we will be implementing in the 2025 field season. As that part of the project evolved, the Data Processes Working Group identified the opportunity and need for gaining efficiencies by redesigning and consolidating the Oracle objects (tables and materialized views) that have historically served these data. The Index Computation Working Group also identified an opportunity to gain efficiencies by consolidating the various scripts that were developed independently by both survey region groups into a workflow that was more accessible and documented.

The Index Computation Working Group developed the gapindex R package, a code repository that consolidates the code that calculates the various standard GAP products (e.g., CPUE, total biomass, size/age composition) for both the Bering Sea and AIGOA survey regions. The Data Processes Working Group was responsible for compiling the data structures needed to support data product tables that were consistent across all of the AFSC GAP survey regions as well as the creation of the GAP\_PRODUCTS oracle schema that will house these consolidated products in the future.

This gap\_products GitHub repository houses the code that will conduct the “standard production run” that produces the new data tables via the gapindex R package and upload those tables to the GAP\_PRODUCTS Oracle schema.

## **Major Advantages**

- Consolidated production tables include all standard data products for all surveys. Data will be provided in the same format, with the same units, and created using the same mathematical methodology. This should limit data pulls, reduce complexity for data access, and reduce complicated secondary data wrangling.
- Consistent naming conventions for schemata, tables, and column metadata. Columns across all tables will use the same naming conventions, units, and data types. Restricting standard data product table content to absolutely necessary columns.

### *Major Advantages*

- Removal of redundant data columns that can be acquired by joining to reference tables is key for providing consistent and up-to-date data while limiting data table sizes.
- Consolidation and repurposing of Oracle schemata. This will help the GAP team limit unnecessary access to unprocessed or problematic data by outside users.
- Vetted data methods. All code and data inclusion decisions and wrangling are documented in the {gapindex} R package. Streamlined and rapid data production. Improved and consolidated data creation and documentation provide data creators and users with greater confidence in the data products and enhanced ability to share the data.

# **1. Survey Background**

## **1.1. What is the research objective?**

The objectives of these surveys are to:

- monitor trends in the marine ecosystem of the Bering Sea, Aleutian Islands, and Gulf of Alaska
- produce fishery-independent biomass and abundance estimates for commercially important fish and crab species
- collect biological and environmental data for use in ecosystem-based fishery management.

Learn more about the program

## **1.2. Who is conducting the research?**

Scientists from the Alaska Fisheries Science Center conduct these bottom trawl surveys with participation from the Alaska Department of Fish & Game (ADF&G), the International Pacific Halibut Commission (IPHC), and universities. This research is conducted on chartered fishing vessels.

## 1. Survey Background

### 1.3. Bottom trawl surveys and regions

## Bottom Trawl Survey Regions

AFSC RACE Groundfish and Shellfish Public Data Coverage



Each survey conducted by the Groundfish Assessment Program are multispecies bottom trawl surveys. We collect environmental and biological data to assess how climate variability and loss of sea ice are affecting bottom-dwelling marine life on the Bering Sea shelf. We monitor trends in the distribution (location and movement patterns) and abundance of groundfish and crab species as well as oceanographic data (e.g., water temperature, depth). We collect biological information such as organism weight, length, stomachs to learn about diets, and otoliths to determine fish ages. We use this information in annual stock assessments and to assess the state of the ecosystem. This research is conducted on fishing industry contract vessels.

Table 1.1.: Survey summary stats

Survey	Survey Definition ID	Years	Depth (m)	Area (km <sup>2</sup> )	# Statistical Areas	# Possible Stations
Aleutian Islands Bottom Trawl Survey	52	2022 - 1980 (16)	1 - 500	64,415.0	80	1,312
Eastern Bering Sea Slope Bottom Trawl Survey	78	2016 - 2002 (6)	201 - 800	21,134.2	4	

## 1. Survey Background

<b>Survey</b>	<b>Survey Definition ID</b>	<b>Years</b>	<b>Depth (m)</b>	<b>Area (km2)</b>	<b># Statistical Areas</b>	<b># Possible Stations</b>
Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey	98	2023 - 1982 (41)	1 - 200	492,989	29	515
Gulf of Alaska Bottom Trawl Survey	47	2023 - 1984 (18)	1 - 1,000	314,087.4	39	6,939
Northern Bering Sea Crab/Groundfish Survey - Eastern Bering Sea Shelf Survey Extension	143	2022 - 2010 (5)	1 - 100	198,866	4	144

### 1.3.1. Aleutian Islands

(Von Szalay and Raring, 2020)

- Upper Continental Slope of the Aleutian Islands from Unimak Pass to Stalemate Bank
- Triennial (1990s)/Biennial since 2000 in even years, since 1992
- Modified Index-Stratified Random of Successful Stations Survey Design
- Important commercial fish species include Atka mackerel, Pacific ocean perch, walleye pollock, Pacific cod, sablefish, and other rockfish species.

### 1.3.2. Gulf of Alaska

(Von Szalay and Raring, 2018)

- Continental Shelf and Upper Slope of the Gulf of Alaska extending from the Islands of Four Mountains 2,300 km east to Dixon Entrance
- Triennial (1990s)/Biennial since 2001 in odd years, since 1991
- Stratified Random Survey Design
- Important commercial species in the Gulf of Alaska include Pacific ocean perch, walleye pollock, Pacific cod, flatfish, and other rockfish species.

## *1. Survey Background*

### **1.3.3. Eastern Bering Sea Shelf**

(Markowitz et al., 2023)

- The continental shelf of the eastern Bering Sea from the Aleutian Islands to the Bering Strait
- Conducted annually since 1982.
- Uses a stratified systematic sampling survey design with fixed stations at center of 20 x 20 nm grid.
- Similar in design to the northern Bering Sea shelf bottom trawl survey.
- Focus species for the Bering Sea include walleye pollock, Pacific cod, Greenland turbot, yellowfin sole, northern rock sole, red king crab, and snow and Tanner crabs.

### **1.3.4. Northern Bering Sea**

(Markowitz et al., 2023)

- The continental shelf of the northern Bering Sea, including the area north of St. Lawrence Island and Norton Sound
- Biennial/Annual; conducted intermittently since 2010
- Uses a stratified systematic sampling survey design with fixed stations at center of 20 x 20 nm grid.
- Similar in design to the eastern Bering Sea shelf bottom trawl survey.

### **1.3.5. Eastern Bering Sea Upper Continental Slope**

(Hoff, 2016)

- The eastern Bering Sea upper continental slope survey area extends from Unalaska and Akutan Islands to the U.S.-Russian Maritime Boundary at 61° N near the International Date Line (166° E to 180° W) at depths from 200 to 1,200 m
- Conducted intermittently since 2002 (funding dependent)
- Modified Index-Stratified Random of Successful Stations Survey Design
- Focus species for the Bering Sea slope include giant grenadier, Pacific ocean perch, popeye grenadier, walleye pollock, and arrowtooth flounder.

## 1. Survey Background

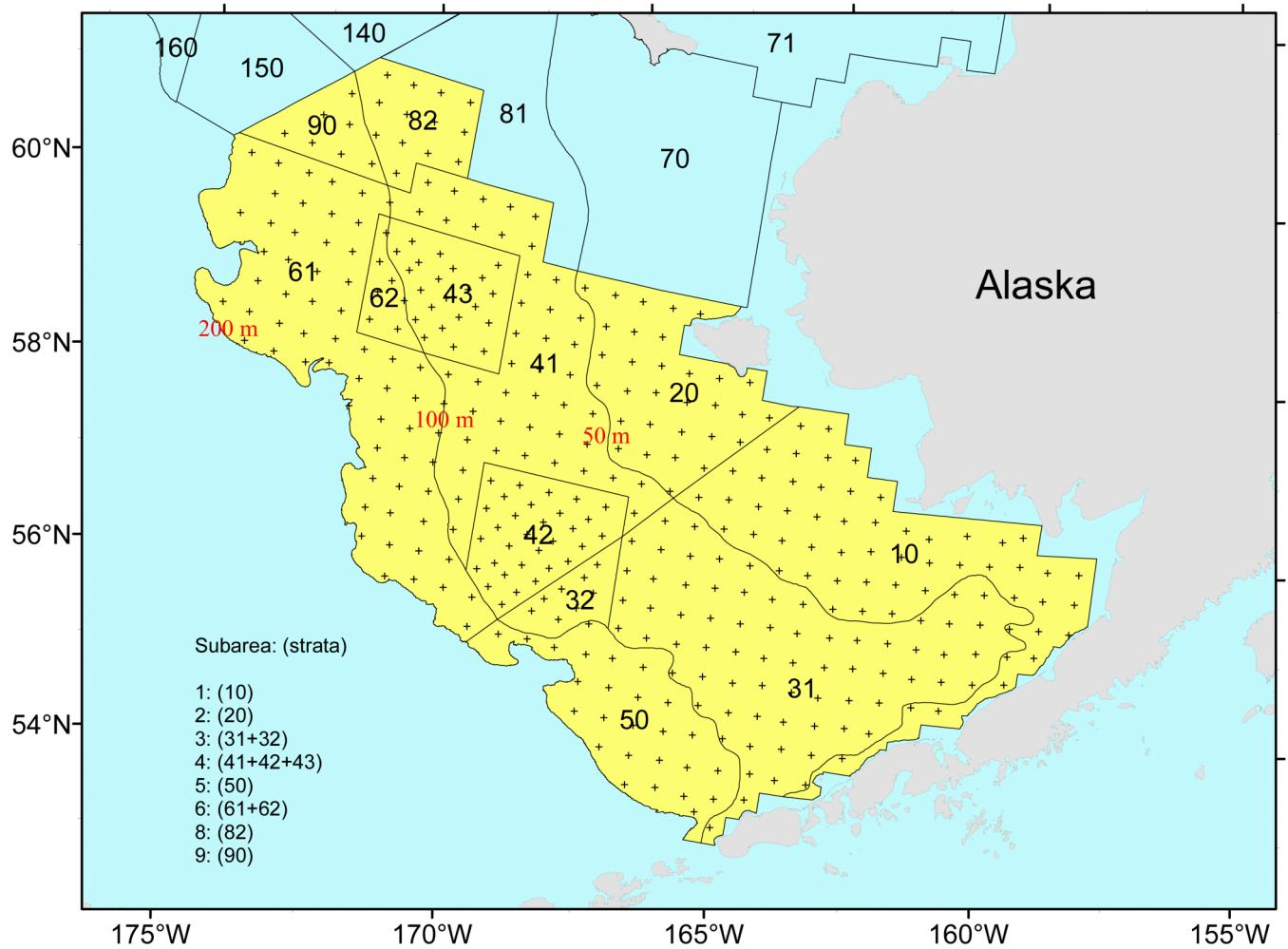


Figure 1.1.: Strata used in the Eastern Bering Sea Survey.

## 1. Survey Background



Figure 1.2.: Strata used in the Bering Sea Slope Survey.

## 2. Workflow

The code/run.R script houses the sequence of programs that calculate the standard data products resulting from the NOAA AFSC GAP bottom trawl surveys. Standard data products are the CPUE, BIOMASS, SIZECOMP, and AGECOMP tables in the GAP\_PRODUCTS Oracle schema. The tables are slated to be updated twice a year, once after the survey season following finalization of that summer's bottom trawl survey data to incorporate the new catch, size, and effort data and once prior to an upcoming survey to incorporate new age data that were processed after the prior summer's survey season ended. This second pre-survey production run will also incorporate changes in the data due to the specimen voucher process as well as other post-hoc changes in the survey data.

Below is a summary of the workflow:

1. Import versions of the tables in GAP\_PRODUCTS locally within the gap\_products repository to compare with the updated production tables. Any changes to a production table will be compared and checked to make sure those changes are intentional and documented.
2. Use the gapindex R package to calculate the four major standard data products: CPUE, BIOMASS, SIZECOMP, AGECOMP. These tables are compared and checked to their respective locally saved copies and any changes to the tables are vetted and documented. These tables are then uploaded to the GAP\_PRODUCTS Oracle schema.
3. Calculate the various materialized views for AKFIN and FOSS purposes. Since these are derivative of the tables in GAP\_PRODUCTS as well as other base tables in RACEBASE and RACE\_DATA, it is not necessary to check these views in addition to the data checks done in the previous steps.

### 2.1. Data levels

GAP produces numerous data products\* that are subjected to different levels of processing, ranging from raw to highly-derived. The suitability of these data products

## *2. Workflow*

for analysis varies and there is ambiguity about which data products can be used for which purpose. This ambiguity can create challenges in communicating about data products and potentially lead to misunderstanding and misuse of data. One approach to communicating about the level of processing applied to data products and their suitability for analysis is to describe data products using a Data Processing Level system. Data Processing Level systems are widely used in earth system sciences to characterize the extent of processing that has been applied to data products. For example, the NOAA National Centers for Environmental Information (NCEI) Satellite Program uses a Data Processing Level system to describe data on a scale of 0-4, where Level 0 is raw data and Level 4 is model output or results from analysis. Example of how NASA remote sensing data products are shared through a public data portal with levels of data processing and documentation.

For more information, see Sean Rohan's October 2022 SCRUGS presentation on the topic.

- **Level 0:** Raw and unprocessed data. Ex: Data on the G drive, some tables in RACE\_DATA
- **Level 1A:** Data products with QA/QC applied that may or may not be expanded to analysis units, but either not georeferenced or does not include full metadata. Ex: Some tables in RACE\_DATA and RACEBASE
- **Level 2:** Analysis-ready data products that are derived for a standardized extent and account for zeros and missing/bad data. Ex: CPUE tables, some data products in public-facing archives and repositories
- **Level 3:** Data products that are synthesized across a standardized extent, often inputs in a higher-level analytical product. Ex: Abundance indices, some data products in public-facing archives and repositories
- **Level 4:** Analytically generated data products that are derived from lower-level data, often to inform management. Ex: Biological reference points from stock assessments, Essential Fish Habitat layers, indicators in Ecosystem Status Reports and Ecosystem and Socioeconomic Profiles

## 3. News

### 3.1. Future plans

#### 3.1.1. GOA 2025 Restratiification – Mock Data for Testing

The plan will be, once all are satisfied with the new GAP\_PRODUCTS schema and tables, to sunset the historic product tables in 2024 and proceed with only GAP\_PRODUCTS for the 2024 post-survey stock assessment season.

- December 2023 - March 2024: Meeting between GAP and stock assessment groups in early December 2023 to update progress on the GAP\_PRODUCTS testing phase. **Deadline for Comments and Feedback on GAP\_PRODUCTS data structures is March 8, 2024.**
- September 2024: GAP will only release data products according to the new standard. Current, historical data product tables will be archived in a new schema called "**GAP\_ARCHIVE**".

### 3.2. Previous updates

- September 2023: Provisional data product tables – CPUE, BIOMASS, SIZECOMP, and AGECOMP – as well as provisional support tables – AREA, STRATUM\_GROUPS, METADATA\_COLUMN, SPECIES\_YEAR, SURVEY\_DESIGN – are available in the GAP\_PRODUCTS Oracle schema with updated 2023 GOA and EBS survey data.
  - Additionally, the inclusion of mock data for the under the new 2025 GOA stratified random survey (labeled in the GAP\_PRODUCTS tables as YEAR 2025) will provide stock authors with the opportunity to interact with data from the new survey design to be implemented in 2025.

### *3. News*

- Provisional AKFIN and FOSS tables are also available in the GAP\_PROD-UCTS Oracle schema. These include: AKFIN\_AGECOMP, AKFIN\_AREA, AKFIN BIOMASS, AKFIN\_CATCH, AKFIN\_CPUE, AKFIN\_CRUISE, AKFIN\_HAUL, AKFIN\_LENGTH, AKFIN\_METADATA\_COLUMN, AKFIN\_SIZECOMP, AKFIN\_SPECIMEN, AKFIN\_SURVEY DESIGN, AKFIN\_STRATUM\_GROUPS, FOSS\_CATCH, FOSS\_CPUE\_PRESONLY, FOSS\_HAUL, and FOSS\_TAXON\_GROUP.
- May 2023: Release of new, draft, standard data product tables, including restratified GOA data. Stock assessment authors will have the opportunity to explore differences between datasets, test workflows, and provide comments and issues during summer 2023.
- February 2023: Decision was made to include the mock restratified GOA data with the development of the new consolidated standard data products.
- December 2022: GAP and SSMA discuss integration of the restratification of the GOA survey design into standard data products.
  - Stock assessors requested a "dry run" test to work with new mock restratified GOA survey data before implementation of the new survey design.
  - This prompted the postponement of the restratified GOA design to 2025.
- October 2022: The data processes and index computation working group convened to address the development of standard survey data products (e.g., biomass/abundance, size composition, age composition, CPUE).
  - Index Computation Working Group: consolidation of index computation methods between the Bering Sea and AI-GOA regions.
  - Data Processes Working Group: consolidation, clean up, and reorganization of survey oracle schemata, tables, and other data for all surveys.

**Part II.**

**GAP Production Data**

## **Data Description**

The Resource Assessment and Conservation Engineering Division (RACE) Groundfish Assessment Program (GAP) of the Alaska Fisheries Science Center (AFSC) conducts fisheries-independent bottom trawl surveys to monitor the condition of the demersal fish and crab stocks of Alaska. These data are developed to describe the temporal distribution and abundance of commercially and ecologically important groundfish species, examine the changes in the species composition of the fauna over time and space, and describe the physical environment of the groundfish habitat.

Users must read and fully comprehend the metadata prior to use. Data should not be used beyond the limits of the source scale. Acknowledgement of NOAA, as the source from which these data were obtained, in any publications and/or other representations of these data, is suggested. These data are compiled and approved annually after each summer survey season. The data from previous years are unlikely to change substantially once published. Some survey data are excluded, such as non-standard stations, surveys completed in earlier years using different/non-standard gear, and special tows and non-standard data collections.

## **Cite this data**

Use the below bibtext citations, as cited in our group's citation repository for citing the data created and maintained in this repo (NOAA Fisheries Alaska Fisheries Science Center, Goundfish Assessment Program, 2023). Add "note = {Accessed: mm/dd/yyyy}" to append the day this data was accessed.

```
@misc{GAPPProducts,
  author = {{NOAA Fisheries Alaska Fisheries Science Center, Goundfish Assessment Program}},
  year = {2023},
  title = {AFSC Goundfish Assessment Program Design-Based Production Data},
  howpublished = {https://www.fisheries.noaa.gov/alaska/science-data/groundfish-assessment-data},
  publisher = {{U.S. Dep. Commer.}},
  copyright = {Public Domain}
}
```

*Data Creation*

## **Data Creation**

These data are created using the gapindex R package.

## **4. Data description**

### **4.1. Data usage**

These tables were created by the Resource Assessment and Conservation Engineering Division (RACE) Groundfish Assessment Program (GAP) of the Alaska Fisheries Science Center (AFSC). There are legal restrictions on access to the data. These data are not intended for public dissemination and should not be shared without the explicit written consent of the data managers and owners (NOAA Fisheries). The GitHub repository for the scripts that created this code can be found at [https://github.com/afsc-gap-products/gap\\_products](https://github.com/afsc-gap-products/gap_products). For more information about codes used in the tables, please refer to the survey code books. These data were last updated September 10, 2023.

### **4.2. Data tables**

#### **4.2.1. AGECOMP**

Region-level age compositions by sex/length bin.

Number of rows: 552856

Number of columns: 9

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AGE

Age bin of taxon

#### *4. Data description*

year

NUMBER(38,0)

Age bin of a taxon in years estimated by the age comp estimate.

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

LENGTH\_MM\_MEAN

Mean length at age weighted by numbers at length

numeric

NUMBER(38,3)

Mean length estimated in age comp estimate.

LENGTH\_MM\_SD

standard deviation of length at age weighted by numbers at length

numeric

NUMBER(38,3)

Variance of mean length estimated in age comp estimate.

POPULATION\_COUNT

Estimated Population

numeric

NUMBER(38,6)

The estimated population caught in the survey for a species, group, or total for a given survey.

SEX

Sex of a specimen

#### *4. Data description*

ID code

NUMBER(38,0)

Sex of a specimen where "1" = "Male", "2" = "Female", "3" = Unsexed.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

YEAR

Year

year

NUMBER(10,0)

Year the survey was conducted in.

#### *4. Data description*

##### **4.2.2. AREA**

[There is currently no description for this table.]

Number of rows: 443

Number of columns: 10

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

AREA\_KM2

Area (km<sup>2</sup>)

kilometers squared

NUMBER(38,3)

Area in thousands of square kilometers.

AREA\_NAME

Area ID Name

text

VARCHAR2(4000 BYTE)

Descriptive name of each AREA\_ID. These names often identify the region, depth ranges, or other regional information for the area ID.

DEPTH\_MAX\_M

#### *4. Data description*

Area ID Maximum Depth (m)

meters

NUMBER(38,3)

Maximum depth (meters) of the area covered by AREA\_ID.

DEPTH\_MIN\_M

Area ID Minimum Depth (m)

meters

NUMBER(38,3)

Minimum depth (meters) of the area covered by AREA\_ID.

DESCRIPTION

Description

text

VARCHAR2(4000 BYTE)

Description of row observation.

DESIGN\_YEAR

Design year

year

NUMBER(10,0)

The year the survey area stratum (e.g., statistical stratum, summary area, region) was implemented in.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

TYPE

#### *4. Data description*

NA

NA

NA

NA

crs

NA

NA

NA

NA

#### **4.2.3. BIOMASS**

Stratum/subarea/region-level mean CPUE (weight and numbers), total biomass, and total abundance with associated variances.

Number of rows: 4589761

Number of columns: 16

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

BIOMASS\_MT

#### *4. Data description*

Estimated Biomass

numeric

NUMBER(38,6)

The estimated biomass caught in the survey for a species, group, or total for a given survey.

BIOMASS\_VAR

Estimated Biomass Variance

numeric

NUMBER(38,6)

The estimated biomass variance caught in the survey for a species, group, or total for a given survey.

CPUE\_KGKM2\_MEAN

Mean Weight CPUE

kilograms per kilometers squared

NUMBER(38,6)

The mean of catch weight (kilograms) divided by area (squared kilometers) swept by the net used in design-based indicie calculation.

CPUE\_KGKM2\_VAR

Variance of the Mean Weight CPUE

kilograms per kilometers squared

NUMBER(38,6)

The variance of mean of catch weight (kilograms) divided by area (squared kilometers) swept by the net used in design-based indicie calculation.

CPUE\_NOKM2\_MEAN

Mean Numeric CPUE

count per kilometers squared

NUMBER(38,6)

The mean of catch count (number) divided by area (squared kilometers) swept by the net used in design-based indicie calculation.

#### *4. Data description*

CPUE\_NOKM2\_VAR

Variance of the Mean Numeric CPUE

count per kilometers squared

NUMBER(38,6)

The variance of mMean of catch count (number) divided by area (squared kilometers) swept by the net used in design-based indicie calculation.

N\_COUNT

Hauls with taxon counts

numeric

NUMBER(38,0)

Total number of hauls with positive taxon counts used in calculation.

N\_HAUL

Valid hauls

numeric

NUMBER(38,0)

Total number of valid hauls used in calculation.

N\_LENGTH

Hauls with taxon lengths

numeric

NUMBER(38,0)

Total number of hauls with taxon length data used in calculation.

N\_WEIGHT

Hauls with catch

numeric

NUMBER(38,0)

Total number of hauls with positive catch/weighed taxon data used in calculation.

POPULATION\_COUNT

Estimated Population

#### *4. Data description*

numeric

NUMBER(38,6)

The estimated population caught in the survey for a species, group, or total for a given survey.

POPULATION\_VAR

Estimated Population Variance

numeric

NUMBER(38,6)

The estimated population variance caught in the survey for a species, group, or total for a given survey.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

YEAR

Year

year

NUMBER(10,0)

Year the survey was conducted in.

#### *4. Data description*

##### **4.2.4. CPUE**

Haul-level zero-filled weight and numerical catch-per-unit-effort.

Number of rows: 37834687

Number of columns: 7

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_SWEPT\_KM2

Area Swept (km)

kilometers

NUMBER(38,6)

The area the net covered while the net was fishing (kilometers squared), defined as the distance fished times the net width.

COUNT

Taxon Count

count, whole number resolution

NUMBER(38,0)

Total number of individuals caught in haul by taxon, represented in whole numbers used in calculation.

CPUE\_KGKM2

Weight CPUE (kg/km<sup>2</sup>)

kilograms per kilometers squared

NUMBER(38,6)

Catch weight (kilograms) divided by area (squared kilometers) swept by the net.

CPUE\_NOKM2

#### *4. Data description*

Number CPUE (no/km<sup>2</sup>)

count per kilometers squared

NUMBER(38,6)

Catch number (in number of organisms) per area (squared kilometers) swept by the net.

HAULJOIN

Haul ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each (vessel, cruise, and haul) combination.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

WEIGHT\_KG

Taxon Weight (kg)

kilograms

NUMBER(38,3)

Weight (thousandths of a kilogram) of individuals in a haul by taxon.

#### *4. Data description*

##### **4.2.5. SURVEY DESIGN**

[There is currently no description for this table.]

Number of rows: 126

Number of columns: 3

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

DESIGN\_YEAR

Design year

year

NUMBER(10,0)

The year the survey area stratum (e.g., statistical stratum, summary area, region) was implemented in.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

YEAR

Year

year

NUMBER(10,0)

Year the survey was conducted in.

#### *4. Data description*

##### **4.2.6. METADATA\_TABLE**

[There is currently no description for this table.]

Number of rows: 8

Number of columns: 3

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

METADATA\_SENTENCE

Sentence

text

VARCHAR2(255 BYTE)

Table metadata sentence.

METADATA\_SENTENCE\_NAME

Metadata sentence name

text

VARCHAR2(255 BYTE)

Name of table metadata sentence.

METADATA\_SENTENCE\_TYPE

Sentence type

text

VARCHAR2(255 BYTE)

Type of sentence to have in table metadata.

#### *4. Data description*

##### **4.2.7. STRATUM\_GROUPS**

[There is currently no description for this table.]

Number of rows: 774

Number of columns: 4

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

DESIGN\_YEAR

Design year

year

NUMBER(10,0)

The year the survey area stratum (e.g., statistical stratum, summary area, region) was implemented in.

STRATUM

Stratum ID

ID code

NUMBER(10,0)

#### *4. Data description*

RACE database statistical area for analyzing data. Strata were designed using bathymetry and other geographic and habitat-related elements. The strata are unique to each survey series. Stratum of value 0 indicates experimental tows.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

#### **4.2.8. SIZECOMP**

[There is currently no description for this table.]

Number of rows: 3130543

Number of columns: 7

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

LENGTH\_MM

Length of a specimen

#### *4. Data description*

millimeters

NUMBER(10,0)

Length of a specimen in millimeters.

POPULATION\_COUNT

Estimated Population

numeric

NUMBER(38,6)

The estimated population caught in the survey for a species, group, or total for a given survey.

SEX

Sex of a specimen

ID code

NUMBER(38,0)

Sex of a specimen where "1" = "Male", "2" = "Female", "3" = Unsexed.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

YEAR

Year

*4. Data description*

year

NUMBER(10,0)

Year the survey was conducted in.

## **5. gapindex R package**

Code to generate design-based catch-per-unit-effort (CPUE), indices of abundance, biomass, and size and age compositions from survey data is available from <https://github.com/afsc-gap-products/gapindex>.

**Part III.**

**AKFIN**

## *The Alaska Fisheries Information Network*

These data are used directly by stock assessors and are provided to The [Alaska Fisheries Information Network (AKFIN)].

## **The Alaska Fisheries Information Network**

The Alaska Fisheries Information Network (AKFIN) is a regional program that consolidates and supports the collection, processing, analysis, and reporting of fisheries statistics for North Pacific and Alaskan fisheries. AKFIN integrates this information into a single data management system using consistent methods and standardized formats. The Network then reports this information on its website, in various publications, and to researchers. The resulting data enables fishery managers, scientists, and associated agencies to supervise fisheries resources more effectively and efficiently.

If you are an AFSC employee with access to data through our internal database Oracle server, use this guide to access our data. If not, reach out to AKFIN for a user account.

## **Cite this data**

Use the below bibtext citations, as cited in our group's citation repository for citing the data created and maintained in this repo (Alaska Fisheries Information Network (AKFIN), 2023). Add "note = {Accessed: mm/dd/yyyy}" to append the day this data was accessed.

```
@misc{GAPakfin,
  author = {{Alaska Fisheries Information Network (AKFIN)}},
  institution = {{NOAA Fisheries Alaska Fisheries Science Center, Goundfish Assessment Program}},
  year = {2023},
  title = {AFSC Goundfish Assessment Program Design-Based Production Data},
  howpublished = {\url{https://www.psmfc.org/program/alaska-fisheries-information-network-akfin}},
  publisher = {{U.S. Dep. Commer.}},
  copyright = {Public Domain}
}
```

# **6. Data description**

## **6.1. Data Description**

**[OUTDATED]** AKFIN Answers <https://akfin.psmfc.org/akfin-answers/> is an Oracle BI tool used for distributing data to stock assessors and other users. Usernames and passwords are distinct from direct akfin database credentials (though they may be identical). RACE data on the AKFIN Answers stock assessment dashboard is located on the “RACE Survey” tab for groundfish and the “Crab” tab for crab surveys. More detailed descriptions of each report are included within that report.

## **6.2. Data Tables**

### **6.3. Data usage**

NA

### **6.4. Data tables**

#### **6.4.1. AKFIN\_AGECOMP**

[There is currently no description for this table.]

Number of rows: 552856

Number of columns: 9

Column name from data

Descriptive column Name

Units

## *6. Data description*

Oracle data type

Column description

AGE

Age bin of taxon

year

NUMBER(38,0)

Age bin of a taxon in years estimated by the age comp estimate.

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

LENGTH\_MM\_MEAN

Mean length at age weighted by numbers at length

numeric

NUMBER(38,3)

Mean length estimated in age comp estimate.

LENGTH\_MM\_SD

standard deviation of length at age weighted by numbers at length

numeric

NUMBER(38,3)

Variance of mean length estimated in age comp estimate.

POPULATION\_COUNT

Estimated Population

numeric

NUMBER(38,6)

## *6. Data description*

The estimated population caught in the survey for a species, group, or total for a given survey.

SEX

Sex of a specimen

ID code

NUMBER(38,0)

Sex of a specimen where "1" = "Male", "2" = "Female", "3" = Unsexed.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

YEAR

Year

year

NUMBER(10,0)

Year the survey was conducted in.

## *6. Data description*

### **6.4.2. AKFIN\_AREA**

[There is currently no description for this table.]

Number of rows: 473

Number of columns: 10

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

AREA\_KM2

Area (km<sup>2</sup>)

kilometers squared

NUMBER(38,3)

Area in thousands of square kilometers.

AREA\_NAME

Area ID Name

text

VARCHAR2(4000 BYTE)

Descriptive name of each AREA\_ID. These names often identify the region, depth ranges, or other regional information for the area ID.

CRS

## *6. Data description*

Coordinate Reference System

ID code

VARCHAR2(5 BYTE)

Coordinate reference system that areas (like AREA\_KM2) are calculated in, as defined by <https://spatialreference.org/> (e.g., "+proj=latlong", "EPSG:3338").

DEPTH\_MAX\_M

Area ID Maximum Depth (m)

meters

NUMBER(38,3)

Maximum depth (meters) of the area covered by AREA\_ID.

DEPTH\_MIN\_M

Area ID Minimum Depth (m)

meters

NUMBER(38,3)

Minimum depth (meters) of the area covered by AREA\_ID.

DESCRIPTION

Description

text

VARCHAR2(4000 BYTE)

Description of row observation.

DESIGN\_YEAR

Design year

year

NUMBER(10,0)

The year the survey area stratum (e.g., statistical stratum, summary area, region) was implemented in.

SURVEY\_DEFINITION\_ID

Survey ID

## *6. Data description*

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

TYPE

NA

NA

NA

NA

### **6.4.3. AKFIN\_BIOMASS**

[There is currently no description for this table.]

Number of rows: 4589761

Number of columns: 16

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

BIOMASS\_MT

## *6. Data description*

Estimated Biomass

numeric

NUMBER(38,6)

The estimated biomass caught in the survey for a species, group, or total for a given survey.

BIOMASS\_VAR

Estimated Biomass Variance

numeric

NUMBER(38,6)

The estimated biomass variance caught in the survey for a species, group, or total for a given survey.

CPUE\_KGKM2\_MEAN

Mean Weight CPUE

kilograms per kilometers squared

NUMBER(38,6)

The mean of catch weight (kilograms) divided by area (squared kilometers) swept by the net used in design-based indicie calculation.

CPUE\_KGKM2\_VAR

Variance of the Mean Weight CPUE

kilograms per kilometers squared

NUMBER(38,6)

The variance of mean of catch weight (kilograms) divided by area (squared kilometers) swept by the net used in design-based indicie calculation.

CPUE\_NOKM2\_MEAN

Mean Numeric CPUE

count per kilometers squared

NUMBER(38,6)

The mean of catch count (number) divided by area (squared kilometers) swept by the net used in design-based indicie calculation.

## *6. Data description*

CPUE\_NOKM2\_VAR

Variance of the Mean Numeric CPUE

count per kilometers squared

NUMBER(38,6)

The variance of mMean of catch count (number) divided by area (squared kilometers) swept by the net used in design-based indicie calculation.

N\_COUNT

Hauls with taxon counts

numeric

NUMBER(38,0)

Total number of hauls with positive taxon counts used in calculation.

N\_HAUL

Valid hauls

numeric

NUMBER(38,0)

Total number of valid hauls used in calculation.

N\_LENGTH

Hauls with taxon lengths

numeric

NUMBER(38,0)

Total number of hauls with taxon length data used in calculation.

N\_WEIGHT

Hauls with catch

numeric

NUMBER(38,0)

Total number of hauls with positive catch/weighed taxon data used in calculation.

POPULATION\_COUNT

Estimated Population

## *6. Data description*

numeric

NUMBER(38,6)

The estimated population caught in the survey for a species, group, or total for a given survey.

POPULATION\_VAR

Estimated Population Variance

numeric

NUMBER(38,6)

The estimated population variance caught in the survey for a species, group, or total for a given survey.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

YEAR

Year

year

NUMBER(10,0)

Year the survey was conducted in.

## *6. Data description*

### **6.4.4. AKFIN\_CATCH**

[There is currently no description for this table.]

Number of rows: 985442

Number of columns: 6

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

CATCHJOIN

NA

NA

NA

NA

COUNT

Taxon Count

count, whole number resolution

NUMBER(38,0)

Total number of individuals caught in haul by taxon, represented in whole numbers used in calculation.

CRUISEJOIN

Cruise ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each survey, vessel, and year combination.

HAULJOIN

## *6. Data description*

Haul ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each (vessel, cruise, and haul) combination.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

WEIGHT\_KG

Taxon Weight (kg)

kilograms

NUMBER(38,3)

Weight (thousandths of a kilogram) of individuals in a haul by taxon.

### **6.4.5. AKFIN\_CPUE**

[There is currently no description for this table.]

Number of rows: 37834687

Number of columns: 7

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_SWEPT\_KM2

Area Swept (km)

## *6. Data description*

kilometers

NUMBER(38,6)

The area the net covered while the net was fishing (kilometers squared), defined as the distance fished times the net width.

COUNT

Taxon Count

count, whole number resolution

NUMBER(38,0)

Total number of individuals caught in haul by taxon, represented in whole numbers used in calculation.

CPUE\_KGKM2

Weight CPUE (kg/km<sup>2</sup>)

kilograms per kilometers squared

NUMBER(38,6)

Catch weight (kilograms) divided by area (squared kilometers) swept by the net.

CPUE\_NOKM2

Number CPUE (no/km<sup>2</sup>)

count per kilometers squared

NUMBER(38,6)

Catch number (in number of organisms) per area (squared kilometers) swept by the net.

HAULJOIN

Haul ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each (vessel, cruise, and haul) combination.

SPECIES\_CODE

Taxon Code

## *6. Data description*

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

WEIGHT\_KG

Taxon Weight (kg)

kilograms

NUMBER(38,3)

Weight (thousandths of a kilogram) of individuals in a haul by taxon.

### **6.4.6. AKFIN\_CRUISE**

[There is currently no description for this table.]

Number of rows: 185

Number of columns: 10

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

CRUISE

Cruise ID

ID code

NUMBER(38,0)

This is a six-digit number identifying the cruise number of the form: YYYY99 (where YYYY = year of the cruise; 99 = 2-digit number and is sequential; 01 denotes the first cruise that vessel made in this year, 02 is the second, etc.).

CRUISEJOIN

Cruise ID

## *6. Data description*

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each survey, vessel, and year combination.

DATE\_END

End Date

YYYY-MM-DD

DATE

The date (YYYY-MM-DD) of the end of the event (e.g., cruise).

DATE\_START

Start Date

YYYY-MM-DD

DATE

The date (YYYY-MM-DD) of the beginning of the event (e.g., cruise).

SPONSOR\_ACRONYM

NA

NA

NA

NA

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

SURVEY\_NAME

NA

## *6. Data description*

NA

NA

NA

VESSEL\_ID

Vessel ID

ID code

NUMBER(38,0)

ID number of the vessel used to collect data for that haul. The column 'vessel\_id' is associated with the 'vessel\_name' column. Note that it is possible for a vessel to have a new name but the same vessel id number. For a complete list of vessel ID codes, review the code books.

VESSEL\_NAME

Vessel Name

text

VARCHAR2(255 BYTE)

Name of the vessel used to collect data for that haul. The column 'vessel\_name' is associated with the 'vessel\_id' column. Note that it is possible for a vessel to have a new name but the same vessel id number. For a complete list of vessel ID codes, review the code books.

YEAR

Year

year

NUMBER(10,0)

Year the survey was conducted in.

## *6. Data description*

### **6.4.7. AKFIN\_HAUL**

[There is currently no description for this table.]

Number of rows: 35998

Number of columns: 25

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

ACCESSORIES

NA

NA

NA

NA

BOTTOM\_TYPE

NA

NA

NA

NA

CRUISEJOIN

Cruise ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each survey, vessel, and year combination.

DATE\_TIME\_START

Start Date and Time

## *6. Data description*

MM/DD/YYYY HH::MM

DATE

The date (MM/DD/YYYY) and time (HH:MM) of the beginning of the haul.

DEPTH\_GEAR\_M

NA

NA

NA

NA

DEPTH\_M

Depth (m)

degrees Celsius

NUMBER(38,1)

Bottom depth (tenths of a meter).

DISTANCE\_FISHED\_KM

Distance Fished (km)

degrees Celsius

NUMBER(38,3)

Distance the net fished (thousandths of kilometers).

DURATION\_HR

Tow Duration (decimal hr)

hours

NUMBER(38,1)

This is the elapsed time between start and end of a haul (decimal hours).

GEAR

NA

NA

NA

## *6. Data description*

NA

GEAR\_TEMPERATURE\_C

NA

NA

NA

NA

HAUL

Haul Number

ID code

NUMBER(38,0)

This number uniquely identifies a sampling event (haul) within a cruise. It is a sequential number, in chronological order of occurrence.

HAULJOIN

Haul ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each (vessel, cruise, and haul) combination.

HAUL\_TYPE

Haul Sampling Type

ID code

NUMBER(38,0)

Type of haul sampling method. For a complete list of haul type ID codes, review the code books.

LATITUDE\_DD\_END

End Latitude (decimal degrees)

decimal degrees

NUMBER(38,6)

## *6. Data description*

Latitude (one hundred thousandth of a decimal degree) of the end of the haul.

LATITUDE\_DD\_START

Start Latitude (decimal degrees)

decimal degrees

NUMBER(38,6)

Latitude (one hundred thousandth of a decimal degree) of the start of the haul.

LONGITUDE\_DD\_END

End Longitude (decimal degrees)

decimal degrees

NUMBER(38,6)

Longitude (one hundred thousandth of a decimal degree) of the end of the haul.

LONGITUDE\_DD\_START

Start Longitude (decimal degrees)

decimal degrees

NUMBER(38,6)

Longitude (one hundred thousandth of a decimal degree) of the start of the haul.

NET\_HEIGHT\_M

Net Height (m)

meters

NUMBER(38,1)

Measured or estimated distance (meters) between footrope and headrope of the trawl.

NET\_MEASURED

Net measured during haul

logical

NA

Logical, describing if the net was measured (TRUE) or not (FALSE) by wheelhouse and marport programs during the haul.

## *6. Data description*

NET\_WIDTH\_M

Net Width (m)

meters

NUMBER(38,1)

Measured or estimated distance (meters) between wingtips of the trawl.

PERFORMANCE

Haul Performance Code

category

NUMBER(38,0)

This denotes what, if any, issues arose during the haul. For more information, review the code books.

STATION

Station ID

ID code

VARCHAR2(255 BYTE)

Alpha-numeric designation for the station established in the design of a survey.

STRATUM

Stratum ID

ID code

NUMBER(10,0)

RACE database statistical area for analyzing data. Strata were designed using bathymetry and other geographic and habitat-related elements. The strata are unique to each survey series. Stratum of value 0 indicates experimental tows.

SURFACE\_TEMPERATURE\_C

Surface Temperature (Degrees Celsius)

degrees Celsius

NUMBER(38,1)

Surface temperature (tenths of a degree Celsius); NA indicates removed or missing values.

## *6. Data description*

WIRE\_LENGTH\_M

Trawl wire length

meters

NUMBER(38,0)

Length of wire deployed during a given haul in meters.

### **6.4.8. AKFIN\_LENGTH**

[There is currently no description for this table.]

Number of rows: 2574443

Number of columns: 7

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

FREQUENCY

NA

NA

NA

NA

HAULJOIN

Haul ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each (vessel, cruise, and haul) combination.

LENGTH\_MM

## *6. Data description*

Length of a specimen

millimeters

NUMBER(10,0)

Length of a specimen in millimeters.

LENGTH\_TYPE

NA

NA

NA

NA

SAMPLE\_TYPE

NA

NA

NA

NA

SEX

Sex of a specimen

ID code

NUMBER(38,0)

Sex of a specimen where "1" = "Male", "2" = "Female", "3" = Unsexed.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

## *6. Data description*

### **6.4.9. AKFIN\_METADATA\_COLUMN**

[There is currently no description for this table.]

Number of rows: 134

Number of columns: 5

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

METADATA\_COLNAME

Column name

text

VARCHAR2(255 BYTE)

Name of the column in a table.

METADATA\_COLNAME\_DESC

column description

text

VARCHAR2(4000 BYTE)

Descripttion of the column.

METADATA\_COLNAME\_LONG

Column name spelled out

text

VARCHAR2(255 BYTE)

Long name for the column.

METADATA\_DATATYPE

NA

NA

## *6. Data description*

NA  
NA  
METADATA\_UNITS  
Units  
category  
VARCHAR2(255 BYTE)  
Units of the column.

### **6.4.10. AKFIN\_SIZECOMP**

[There is currently no description for this table.]

Number of rows: 3130543

Number of columns: 7

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

LENGTH\_MM

Length of a specimen

millimeters

NUMBER(10,0)

## *6. Data description*

Length of a specimen in millimeters.

POPULATION\_COUNT

Estimated Population

numeric

NUMBER(38,6)

The estimated population caught in the survey for a species, group, or total for a given survey.

SEX

Sex of a specimen

ID code

NUMBER(38,0)

Sex of a specimen where "1" = "Male", "2" = "Female", "3" = Unsexed.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

YEAR

Year

year

NUMBER(10,0)

## *6. Data description*

Year the survey was conducted in.

### **6.4.11. AKFIN\_SPECIMEN**

[There is currently no description for this table.]

Number of rows: 359317

Number of columns: 14

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AGE\_DETERMINATION\_METHOD

NA

NA

NA

NA

AGE\_YEARS

NA

NA

NA

NA

CRUISEJOIN

Cruise ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each survey, vessel, and year combination.

## *6. Data description*

GONAD\_G

NA

NA

NA

NA

HAULJOIN

Haul ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each (vessel, cruise, and haul) combination.

LENGTH\_MM

Length of a specimen

millimeters

NUMBER(10,0)

Length of a specimen in millimeters.

MATURITY

NA

NA

NA

NA

MATURITY\_TABLE

NA

NA

NA

NA

SEX

Sex of a specimen

## *6. Data description*

ID code

NUMBER(38,0)

Sex of a specimen where "1" = "Male", "2" = "Female", "3" = Unsexed.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

SPECIMEN\_ID

NA

NA

NA

NA

SPECIMEN\_SAMPLE\_TYPE

NA

NA

NA

NA

SPECIMEN\_SUBSAMPLE\_METHOD

NA

NA

NA

NA

WEIGHT\_G

NA

NA

NA

## *6. Data description*

NA

### **6.4.12. AKFIN\_STRATUM\_GROUPS**

[There is currently no description for this table.]

Number of rows: 774

Number of columns: 4

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_ID

Area ID code

ID code

NUMBER(38,0)

Area ID code for each statistical area used to produce production estimates (e.g., biomass, population, age comps, length comps). Each area ID is unique within each survey.

DESIGN\_YEAR

Design year

year

NUMBER(10,0)

The year the survey area stratum (e.g., statistical stratum, summary area, region) was implemented in.

STRATUM

Stratum ID

ID code

NUMBER(10,0)

## *6. Data description*

RACE database statistical area for analyzing data. Strata were designed using bathymetry and other geographic and habitat-related elements. The strata are unique to each survey series. Stratum of value 0 indicates experimental tows.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

### **6.4.13. AKFIN\_SURVEY\_DESIGN**

[There is currently no description for this table.]

Number of rows: 126

Number of columns: 4

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

DESIGN\_YEAR

Design year

year

NUMBER(10,0)

The year the survey area stratum (e.g., statistical stratum, summary area, region) was implemented in.

SURVEY

Survey Name

text

## *6. Data description*

VARCHAR2(255 BYTE)

Name and description of survey. The column 'survey' is associated with the 'srvy' and 'survey\_id' columns.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

YEAR

Year

year

NUMBER(10,0)

Year the survey was conducted in.

### **6.4.14. AKFIN\_TAXONOMY**

[There is currently no description for this table.]

Number of rows: 2757

Number of columns: 19

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

CLASS\_TAXON

Class phylogenetic rank

category

## *6. Data description*

VARCHAR2(255 BYTE)

Phylogenetic latin rank of class\_taxon of a given species.

COMMON\_NAME

Taxon Common Name

text

VARCHAR2(255 BYTE)

The common name of the marine organism associated with the 'scientific\_name' and 'species\_code' columns. For a complete species list, review the code books.

DATABASE

Genus phylogenetic rank

category

VARCHAR2(255 BYTE)

Taxonomic database source (e.g., "ITIS", "WORMS").

DATABASE\_ID

Subfamily phylogenetic rank

ID code

VARCHAR2(255 BYTE)

Species ID code of a species in the taxonomic "DATABASE" source.

FAMILY\_TAXON

NA

NA

NA

NA

GENUS\_TAXON

Genus phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of genus\_taxon of a given species.

## *6. Data description*

ID\_RANK

Lowest taxonomic rank

text

VARCHAR2(255 BYTE)

Lowest taxonomic rank of a given species entry.

KINGDOM\_TAXON

kingdom\_taxon phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of kingdom\_taxon of a given species.

ORDER\_TAXON

Subclass phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of order\_taxon of a given species.

PHYLUM\_TAXON

phylum\_taxon phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of phylum\_taxon of a given species.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

SPECIES\_NAME

NA

## *6. Data description*

NA

NA

NA

SUBCLASS\_TAXON

Subclass phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of subclass\_taxon of a given species.

SUBFAMILY\_TAXON

Subfamily phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of subfamily\_taxon of a given species.

SUBORDER\_TAXON

Suborder phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of suborder\_taxon of a given species.

SUBPHYLUM\_TAXON

Subphylum phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of subphylum\_taxon of a given species.

SUPERCLASS\_TAXON

Superclass phylogenetic rank

category

VARCHAR2(255 BYTE)

## *6. Data description*

Phylogenetic latin rank of superclass\_taxon of a given species.

SUPERFAMILY\_TAXON

Superfamily phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of superfamily\_taxon of a given species.

SUPERORDER\_TAXON

NA

NA

NA

NA

# **7. Accessing Data**

## **7.1. Access data via Oracle (AFSC only)**

AFSC Oracle users can access the database via SQL developer to view and pull the production data directly from the GAP\_PRODUCTS Oracle schema. The user can also use SQL developer to view and pull the GAP Products data directly from the GAP\_PRODUCTS Oracle schema.

### **7.1.1. Connect to Oracle from R**

Many users will want to access the data from Oracle using R. The user will need to install the RODBC R package and ask OFIS (IT) connect R to Oracle. Then, use the following code in R to establish a connection from R to Oracle:

Here, the user can establish the oracle connection by entering their username and password in the channel `<- gapindex::oracle_connect()` function. Never save usernames or passwords in scripts that may be intentionally or unintentionally shared with others. If no username and password is entered in the function, pop-ups will appear on the screen asking for the username and password.

## **7.2. Data SQL Query Examples:**

### **7.2.1. Ex. 0: Select all data from a table**

You can download all of the tables locally using a variation of the code below. Once connected, pull and save the tables of interest into the R environment.

## 7. Accessing Data

```
locations <- c(
  "GAP_PRODUCTS.AKFIN_AGECOMP",
  "GAP_PRODUCTS.AKFIN_AREA",
  "GAP_PRODUCTS.AKFIN_BIOMASS",
  "GAP_PRODUCTS.AKFIN_CATCH",
  "GAP_PRODUCTS.AKFIN_CPUE",
  "GAP_PRODUCTS.AKFIN_CRUISE",
  "GAP_PRODUCTS.AKFIN_HAUL",
  "GAP_PRODUCTS.AKFIN_LENGTH",
  "GAP_PRODUCTS.AKFIN_METADATA_COLUMN",
  "GAP_PRODUCTS.AKFIN_SIZECOMP",
  "GAP_PRODUCTS.AKFIN_SPECIMEN",
  "GAP_PRODUCTS.AKFIN_STRATUM_GROUPS",
  "GAP_PRODUCTS.AKFIN_SURVEY_DESIGN",
  "GAP_PRODUCTS.AKFIN_TAXONOMY"
)

for (i in 1:length(locations)) {
  print(locations[i])
  a <- RODBC::sqlQuery(channel, paste0("SELECT * FROM ", locations[i]))
  write.csv(x = a, file = here::here("data", paste0(locations[i], ".csv")))
}
```

### 7.2.2. Ex. 1: GOA Pacific Ocean perch biomass and abundance

Biomass and abundance for Pacific Ocean perch from 1990 – 2023 for the western/central/eastern GOA management areas as well as for the entire region.

```
dat <- RODBC::sqlQuery(channel = channel,
                        query =
"WITH FILTERED_STRATA AS (
SELECT AREA_ID, DESCRIPTION FROM GAP_PRODUCTS.AKFIN_AREA
WHERE TYPE in ('REGULATORY_AREA', 'REGION')
AND SURVEY_DEFINITION_ID = 47)
SELECT
BIOMASS_MT,
POPULATION_COUNT,
YEAR,
DESCRIPTION
```

## 7. Accessing Data

```
FROM GAP_PRODUCTS.AKFIN_BIOMASS BIOMASS
JOIN FILTERED_STRATA STRATA
ON STRATA.AREA_ID = BIOMASS.AREA_ID
WHERE BIOMASS.SURVEY_DEFINITION_ID IN 47
AND BIOMASS.SPECIES_CODE = 30060")
```

```
dat0 <- dat %>%
  janitor::clean_names() %>%
  dplyr::select(biomass_mt, population_count, year, area = description) %>%
  pivot_longer(cols = c("biomass_mt", "population_count"),
               names_to = "var",
               values_to = "val") %>%
  dplyr::mutate(
    val = ifelse(var == "biomass_mt", val/1e6, val/1e9),
    var = ifelse(var == "biomass_mt", "Biomass (Mmt)", "Population (B)"),
    area = gsub(x = area, pattern = " - ", replacement = "\n"),
    area = gsub(x = area, pattern = ": ", replacement = "\n"),
    type = sapply(X = strsplit(x = area, split = "\n", fixed = TRUE), `[[`, 2)) %>%
  dplyr::arrange(type) %>%
  dplyr::mutate(
    area = factor(area, levels = unique(area), labels = unique(area), ordered = TRUE))

flextable::flextable(head(dat)) %>%
  theme_zebra() %>%
  flextable::colformat_num(x = ., j = "YEAR", big.mark = "")
```

Table 7.1.: Ex. 1: GOA Pacific Ocean perch biomass and abundance.

<b>BIOMASS_POPULATI MT COUNT</b>	<b>YEARDESCRIPTION</b>
483,622.6833,902,16	GOA 1993Region: All Strata
483,622.6833,902,161	GOA 1993Region: All Strata

## 7. Accessing Data

BIOMASS_POPULATI MT COUNT	YEARDESCRIPTION
771,412.81,252,616,600	GOA 1996Region: All Strata
771,412.81,252,616,603	GOA 1996Region: All Strata
727,063.51,212,034,913	GOA 1999Region: All Strata
727,063.51,212,034,913	GOA 1999Region: All Strata

```
# install.packages("scales")
library(scales)
figure <- ggplot2::ggplot(
  dat = dat0,
  mapping = aes(x = year, y = val, color = type)) +
  ggplot2::geom_point(size = 3) +
  ggplot2::facet_grid(cols = vars(area), rows = vars(var), scales = "free_y") +
  ggplot2::scale_x_continuous(name = "Year", n.breaks = 3) +
  ggplot2::scale_y_continuous(name = "Estimate", labels = comma) +
  ggplot2::labs(title = 'GOA Pacific Ocean perch biomass and abundance 1990 - 2023') +
  ggplot2::guides(color=guide_legend(title = "Region Type"))+
  ggplot2::scale_color_grey() +
  ggplot2::theme_bw() +
  ggplot2::theme(legend.direction = "horizontal",
                legend.position = "bottom")

figure
```

## 7. Accessing Data

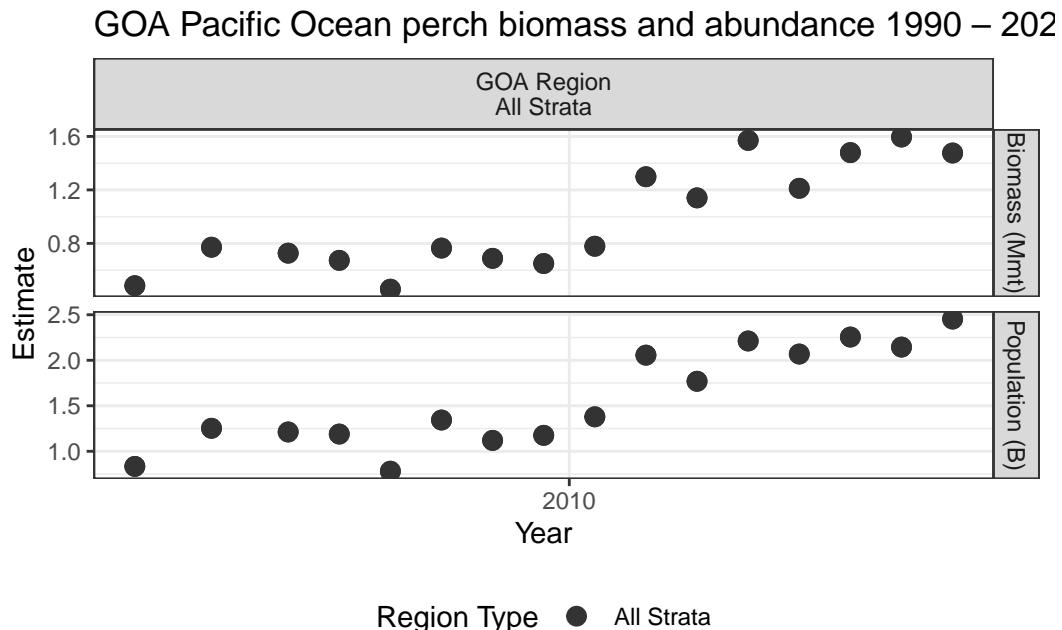


Figure 7.1.: Ex. 1: GOA Pacific Ocean perch biomass and abundance.

### 7.2.3. Ex. 2: AI Rock sole size compositions and ridge plot

Northern and Southern rock sole size composition data from 1991 – 2022 for the Aleutian Islands, with Ridge plot from `ggridges`.

```
dat <- RODBC::sqlQuery(channel = channel,
                        query =
"WITH FILTERED_STRATA AS (
SELECT
AREA_ID,
DESCRIPTION
FROM GAP_PRODUCTS.AKFIN_AREA
WHERE TYPE = 'REGION'
AND SURVEY_DEFINITION_ID = 52)
SELECT
LENGTH_MM,
YEAR
FROM GAP_PRODUCTS.AKFIN_SIZECOMP SIZECOMP
```

## 7. Accessing Data

```

JOIN FILTERED_STRATA STRATA
ON STRATA.AREA_ID = SIZECOMP.AREA_ID
WHERE SIZECOMP.SURVEY_DEFINITION_ID IN 52
AND SIZECOMP.SPECIES_CODE IN (10261, 10262)")

```

```

dat0 <- dat %>%
  janitor::clean_names() %>%
  dplyr::mutate(length_cm = length_mm/10)
flextable::flextable(head(dat)) %>%
  theme_zebra() %>%
  flextable::colformat_num(x = ., j = "YEAR", big.mark = "")

```

Table 7.2.: Ex. 2: AI Rock sole size compositions and ridge plot.

LENGTH-MM	YEAR
160	2000
170	2000
180	2000
190	2000
200	2000
210	2000

```

# install.packages("ggridges")
library(ggridges)
figure <-
  ggplot2::ggplot(
    data = dat0,
    mapping = aes(x = length_cm, y = as.factor(year), fill = stat(x))) +
  ggridges::theme_ridges(center_axis_labels = TRUE) +
  ggridges::geom_density_ridges_gradient(scale = 4, show.legend = FALSE) +
  ggplot2::scale_y_discrete(name = "Year", expand = c(0.01, 0)) +
  ggplot2::scale_x_continuous(name = "Length (cm)", expand = c(0.01, 0)) +
  # ggplot2::scale_fill_grey() +
  ggplot2::labs(title = 'AI Rock sole Size Compositions 1991 - 2022')

figure

```

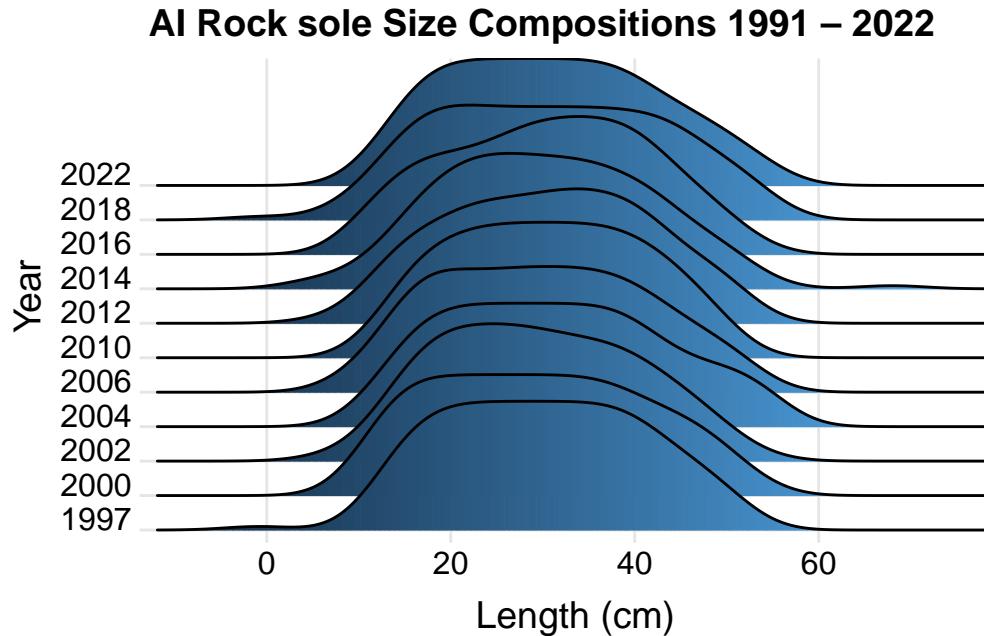


Figure 7.2.: Ex. 2: AI Rock sole size compositions and ridge plot.

#### 7.2.4. Ex. 3: EBS Walleye Pollock Age Compositions and Age Pyramid

Walleye pollock age composition for the EBS Standard Area from 1982 – 2022 and the EBS + NW Area from 1987 – 2022, with age pyramid plot.

```
dat <- RODBC::sqlQuery(channel = channel,
                        query =
"WITH FILTERED_STRATA AS (
SELECT
AREA_ID,
DESCRIPTION
FROM GAP_PRODUCTS.AKFIN_AREA
WHERE TYPE = 'REGION' AND
SURVEY_DEFINITION_ID = 98)
SELECT
AGECOMP.AGE,
```

## 7. Accessing Data

```

AGECOMP.POPULATION_COUNT,
AGECOMP.SEX
FROM GAP_PRODUCTS.AKFIN_AGECOMP AGECOMP
JOIN FILTERED_STRATA STRATA
ON STRATA.AREA_ID = AGECOMP.AREA_ID
WHERE SURVEY_DEFINITION_ID = 98
AND SPECIES_CODE = 21740
AND AGE >= 0")

```

```

dat0 <- dat %>%
  janitor::clean_names() %>%
  dplyr::filter(sex %in% c(1,2)) %>%
  dplyr::mutate(
    sex = ifelse(sex == 1, "M", "F"),
    population_count = # change male population to negative
      ifelse(sex=="M", population_count*(-1), population_count*1)/1e9)

flextable::flextable(head(dat)) %>% theme_zebra()

```

Table 7.3.: Ex. 3: EBS Walleye Pollock Age Compositions and Age Pyramid.

AGE	POPULATI COUNT	SEX
	6446,303,56	1
	7312,858,766	1
	8217,910,31	1
	9145,857,926	1
	1065,496,187	1
	1148,249,148	1

```

figure <- ggplot2::ggplot(
  data = dat0,
  mapping =
    aes(x = age,
        y = population_count,
        fill = sex)) +

```

## 7. Accessing Data

```
ggplot2::scale_fill_grey() +  
  ggplot2::geom_bar(stat = "identity") +  
  ggplot2::coord_flip() +  
  ggplot2::scale_x_continuous(name = "Age") +  
  ggplot2::scale_y_continuous(name = "Population (billions)", labels = abs) +  
  ggplot2::ggttitle(label = "EBS Walleye Pollock Age Compositions 1982 – 2022") +  
  ggplot2::guides(fill = guide_legend(title = "Sex"))+  
  ggplot2::theme_bw()  
  
figure
```

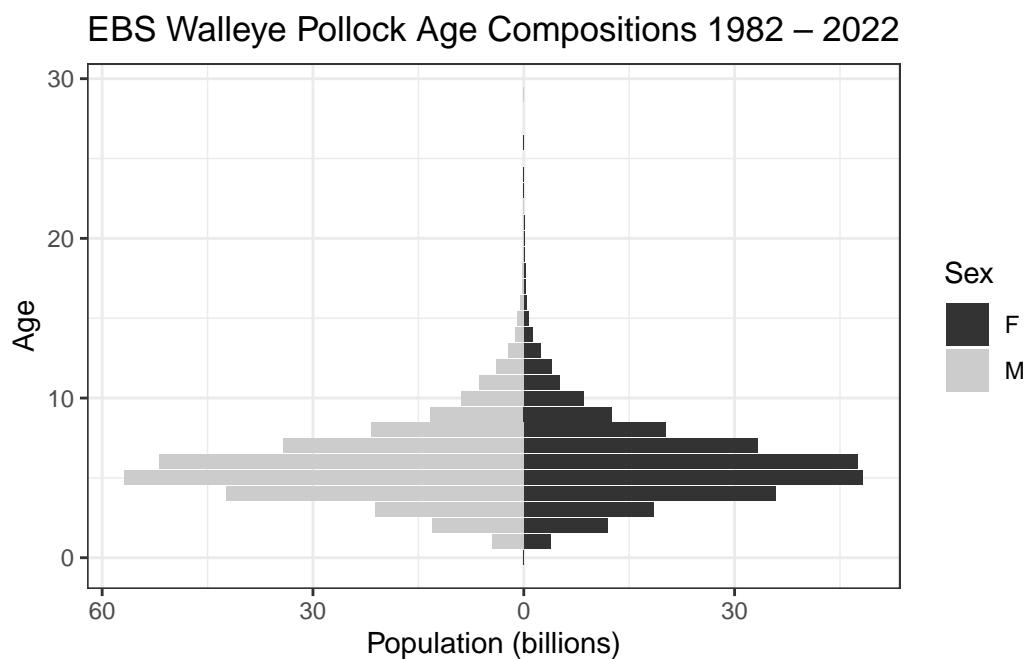


Figure 7.3.: Ex. 3: EBS Walleye Pollock Age Compositions and Age Pyramid.

### 7.2.5. Ex. 4: NBS Pacific cod biomass and abundance

Pacific cod biomass and abundance data for the NBS by stratum.

```
dat <- RODBC::sqlQuery(channel = channel,  
                        query =
```

## 7. Accessing Data

```
"WITH FILTERED_STRATA AS (
SELECT
AREA_ID,
AREA_NAME,
DESCRIPTION
FROM GAP_PRODUCTS.AKFIN_AREA
WHERE TYPE in ('STRATUM') AND
SURVEY_DEFINITION_ID = 143)
SELECT
BIOMASS.BIOMASS_MT,
BIOMASS.POPULATION_COUNT,
BIOMASS.YEAR,
STRATA.AREA_NAME
FROM GAP_PRODUCTS.AKFIN_BIOMASS BIOMASS
JOIN FILTERED_STRATA STRATA
ON STRATA.AREA_ID = BIOMASS.AREA_ID
WHERE BIOMASS.SURVEY_DEFINITION_ID IN 143
AND BIOMASS.SPECIES_CODE = 21720")
```

```
dat0 <- dat %>%
  janitor::clean_names() %>%
  dplyr::select(biomass_mt, population_count, year, area = area_name) %>%
  pivot_longer(cols = c("biomass_mt", "population_count"),
               names_to = "var",
               values_to = "val") %>%
  dplyr::mutate(
    val = ifelse(var == "biomass_mt", val/1e6, val/1e9),
    var = ifelse(var == "biomass_mt", "Biomass (Mmt)", "Population (B)"),
    area = factor(area, levels = unique(area), labels = unique(area), ordered = TRUE))
flextable::flextable(head(dat)) %>%
  theme_zebra() %>%
  flextable::colformat_num(x = ., j = "YEAR", big.mark = "")
```

## 7. Accessing Data

Table 7.4.: Ex. 4: NBS Pacific cod biomass and abundance.

<b>BIOMASS_POPULATI MT COUNT</b>	<b>YEAR</b>	<b>AREA_- NAME</b>
194,846,773,495,085	2019	Inner Domain
194,846,773,495,085	2019	Inner Domain
132,490,266,187,245	2017	Inner Domain

```
figure <- ggplot2::ggplot(
  dat = dat0,
  mapping = aes(y = val, x = year, fill = area)) +
  ggplot2::geom_bar(position="stack", stat="identity") +
  ggplot2::facet_grid(rows = vars(var), scales = "free_y") +
  ggplot2::scale_y_continuous(name = "Estimate", labels = comma) +
  ggplot2::scale_x_continuous(name = "Year", breaks = unique(dat0$year)) +
  ggplot2::labs(title = 'NBS Pacific cod biomass and abundance by stratum') +
  ggplot2::guides(fill=guide_legend(title = "Region Type"))+
  ggplot2::scale_fill_grey() +
  ggplot2::theme_bw() +
  ggplot2::theme(legend.direction = "horizontal",
                legend.position = "bottom")

figure
```

## 7. Accessing Data

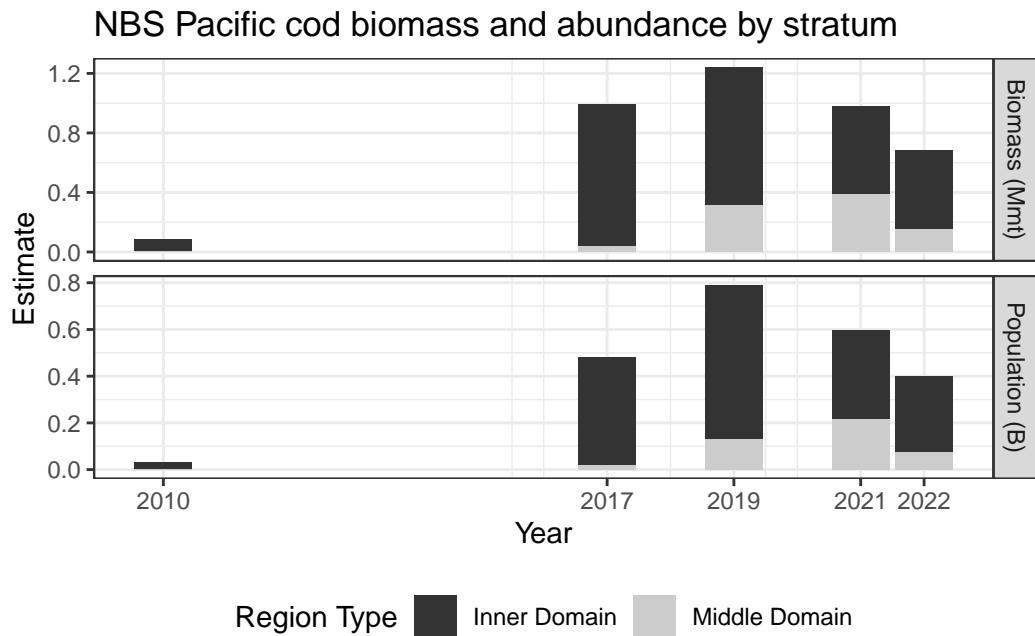


Figure 7.4.: Ex. 4: NBS Pacific cod biomass and abundance.

### 7.2.6. Ex. 5: GOA Pacific Ocean perch biomass and line plot

Pacific Ocean perch biomass totals for GOA between 1984-2021 from GAP\_PRODUCTS.AKFIN\_BIOMASS

```
dat <- RODBC::sqlQuery(channel = channel,
                        query =
"SELECT
SURVEY_DEFINITION_ID,
BIOMASS_MT,
BIOMASS_VAR,
YEAR
FROM GAP_PRODUCTS.AKFIN_BIOMASS
WHERE SPECIES_CODE = 30060
AND SURVEY_DEFINITION_ID = 47
AND AREA_ID = 99903
AND YEAR BETWEEN 1984 AND 2023;") %>%
  janitor::clean_names() %>%
```

## 7. Accessing Data

```
dplyr::mutate(biomass_kmt = biomass_mt/1000,
  # **approximate** 95% confidence interval
  biomass_kci_up = (biomass_mt + (2*sqrt(biomass_var)))/1000,
  biomass_kci_dw = (biomass_mt - (2*sqrt(biomass_var)))/1000)

flextable::flextable(head(dat)) %>%
  theme_zebra() %>%
  flextable::colformat_num(x = ., j = "year", big.mark = "")
```

Table 7.5.: Ex. 5: GOA Pacific Ocean perch biomass and line plot.

survey_- defini- tion_id	biomass_- mt	biomass_- var	year	biomass_- kmt	biomass_- kci_up	biomass_- kci_dw
47	483,622.611,803,384		1993	483.6226	700.9093266.33581	
47	771,412.841,434,152,202		1996	771.41281,178.5204364.30515		
47	727,063.5150,983,54		1999	727.06351,504.1955	-50.06854	
47	673,155.149,285,342,922		2001	673.15511,117.1611229.14901		
47	457,421.65,186,126,!		2003	457.4216	601.4511313.39204	
47	764,901.421,499,807,010		2005	764.90141,058.1577471.64517		

```
a_mean <- dat %>%
  dplyr::group_by(survey_definition_id) %>%
  dplyr::summarise(biomass_kmt = mean(biomass_kmt, na.rm = TRUE),
    minyr = min(year, na.rm = TRUE),
    maxyr = max(year, na.rm = TRUE))

figure <-
  ggplot(data = dat,
    mapping = aes(x = year,
      y = biomass_kmt)) +
  ggplot2::geom_point(size = 2.5, color = "grey40") +
  ggplot2::scale_x_continuous(
    name = "Year",
    labels = scales::label_number(
      accuracy = 1,
```

## 7. Accessing Data

```
    big.mark = ""))
ggplot2::scale_y_continuous(
  name = "Biomass (Kmt)",
  labels = comma) +
ggplot2::geom_segment(
  data = a_mean,
  mapping = aes(x = minyr,
                 xend = maxyr,
                 y = biomass_kmt,
                 yend = biomass_kmt),
  linetype = "dashed",
  linewidth = 2) +
ggplot2::geom_errorbar(
  mapping = aes(ymin = biomass_kci_dw, ymax = biomass_kci_up),
  position = position_dodge(.9),
  alpha = 0.5, width=.2) +
ggplot2::ggttitle(
  label = "GOA Pacific Ocean Perch Biomass 1984-2021",
  subtitle = paste0("Mean = ",
                    formatC(x = a_mean$biomass_kmt,
                           digits = 2,
                           big.mark = ",",
                           format = "f"),
                    " Kmt")) +
ggplot2::theme_bw()

figure
```

## 7. Accessing Data

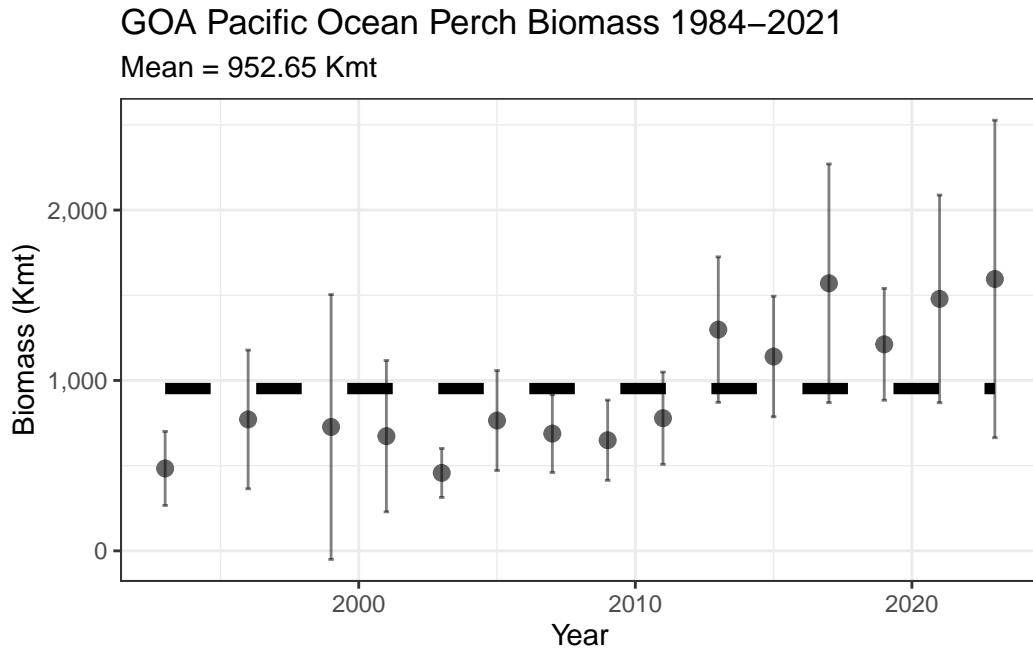


Figure 7.5.: Ex. 5: GOA Pacific Ocean perch biomass and line plot.

### 7.2.7. Ex. 6: EBS Pacific Ocean perch CPUE and akgfmaps map

Pacific Ocean perch catch-per-unit-effort estimates for EBS in 2021 from GAP\_PRODUCTS.AKFIN\_CPUE and map constructed using akgfmaps. Here, we'll use AKFIN HAUL and CRUISES data also included in this repo, for convenience, though they are very similar to their RACEBASE analogs.

```
dat <- RODBC::sqlQuery(channel = channel,
                        query =
"SELECT
(cp.CPUE_KGKM2/100) CPUE_KGHA, -- akgfmaps is expecting hectares
hh.LATITUDE_DD_START LATITUDE,
hh.LONGITUDE_DD_START LONGITUDE

FROM GAP_PRODUCTS.AKFIN_CPUE cp

-- Use HAUL data to obtain LATITUDE & LONGITUDE and connect to cruisejoin
LEFT JOIN GAP_PRODUCTS.AKFIN_HAUL hh
```

## 7. Accessing Data

```

ON cp.HAULJOIN = hh.HAULJOIN

-- Use CRUISES data to obtain YEAR and SURVEY_DEFINITION_ID
LEFT JOIN GAP_PRODUCTS.AKFIN_CRUISE cc
ON hh.CRUISEJOIN = cc.CRUISEJOIN

WHERE cp.SPECIES_CODE = 30060
AND cc.SURVEY_DEFINITION_ID = 98
AND cc.YEAR = 2021;")

```

```
flextable::flextable(head(dat)) %>% theme_zebra()
```

Table 7.6.: Ex. 6: EBS Pacific Ocean perch CPUE and akgfmaps map.

CPUE - KGHA	LATITUDE	LONGITUDE
0	60.65295	-176.2033
0	59.34938	-175.0900
0	61.68338	-173.6652
0	59.97450	-175.9149
0	61.32219	-176.3127
0	61.64331	-175.0828

```

# devtools::install_github("afsc-gap-products/akgfmaps", build_vignettes = TRUE)
library(akgfmaps)

figure <- akgfmaps::make_idw_map(
  x = dat, # Pass data as a data frame
  region = "bs.south", # Predefined EBS area
  set.breaks = "jenks", # Gets Jenks breaks from classint::classIntervals()
  in.crs = "+proj=longlat", # Set input coordinate reference system
  out.crs = "EPSG:3338", # Set output coordinate reference system
  grid.cell = c(20000, 20000), # 20x20km grid
  key.title = "Pacific Ocean perch") # Include in the legend title

[inverse distance weighted interpolation]

```

## 7. Accessing Data

[inverse distance weighted interpolation]

```
figure$plot +  
  ggplot2::guides(fill=guide_legend(title = "Pacific Ocean perch\nCPUE (kg/km2)")) |>  
  change_fill_color(new.scheme = "grey", show.plot = FALSE)
```

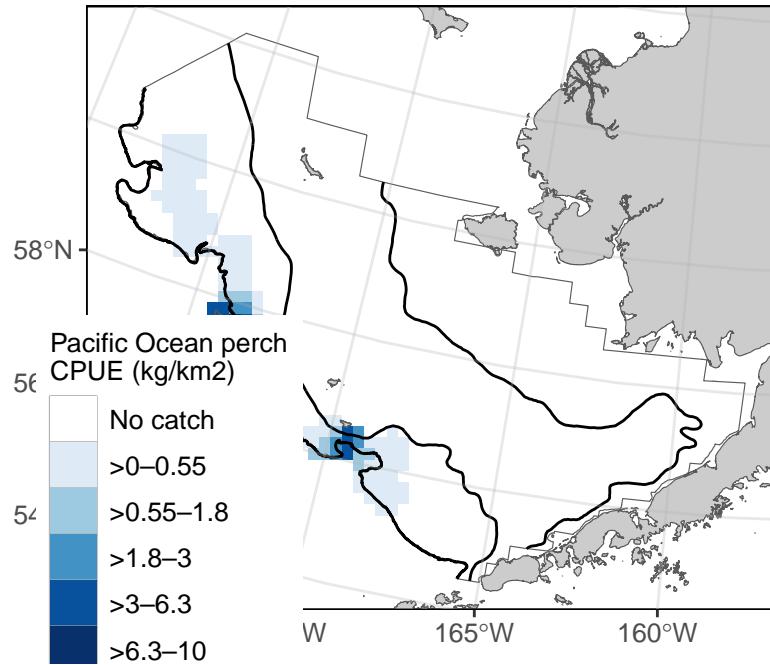


Figure 7.6.: Ex. 6: EBS Pacific Ocean perch CPUE and akgridmaps map.

## 8. Access API data using R

Use the below function to pull AKFIN data through AKFIN's API.

```
# load libraries
library(dplyr)
library(magrittr)
library(httr)
library(flextable)

# tell R to not use scientific notation
options(scipen=999)

# function for pulling data from the api using the httr package
get_gap_biomass<-function(area_id, species_code) {
  # paste(... collapse=",") puts commas between vector elements
  area_id <- paste(area_id, collapse = ",")
  species_code <- paste(species_code, collapse = ",")
  # httr code, parameters are after the '?'
  httr::content(
    httr::GET(paste0("https://apex.psmfc.org/akfin/data_marts/akmp/gap_biomass?area_id=",
                     area_id,
                     "&species_code=",
                     species_code)),
    type = "application/json") %>%
    # convert to data frame
    bind_rows()
}
```

### 8.1. Ex. 1: Load lingcod data

## *8. Access API data using R*

```
lingcod_biomass <- get_gap_biomass(area_id=c(40, 41), species_code=21910)
flextable::flextable(head(lingcod_biomass)) %>%
  flextable::theme_zebra()
```

**Part IV.**

**Public Data (FOSS)**

The final, validated survey data are publicly accessible soon after surveys are completed on the Fisheries One Stop Shop (FOSS) platform. This data includes catch, haul, and environmental data collected at each station. On the FOSS data platform, users can interactively select, view, and download data. Descriptive documentation and user-examples are available on the metadata page.

This data contains all of the catch, environmental, and haul data from the fisheries-independent Groundfish and Shellfish Assessment Program surveys in the Bering Sea, Aleutian Islands, and Gulf of Alaska. This data is sought after by the general public, private entities, and NOAA partners alike, including tribal organizations, K-12 classrooms, academic institutions, for-profit groups, and non-profit groups. This data is compiled and approved once a year after each summer survey season and is available for open access.

**Part V.**

**Collaborators and data users**

### *Cite this data*

Below are a few packages and products currently using this data. If you have developed a product, performed an analysis, or exhibited this data in any way, reach out so we can showcase your hard work.

- **NOAA Fisheries Distribution Mapping and Analysis Portal;** NOAA Fisheries Office of Science and Technology
- **Pull data with python and explore the in-browser visualization tool. Reference their example Python notebook;** The Eric and Wendy Schmidt Center for Data Science and the Environment at UC Berkeley, including sam.pottinger@berkeley.edu, ccmartinez@berkeley.edu, gzarpellon@berkeley.edu, and kkoy@berkeley.edu.

## **Cite this data**

Use the below bibtext citations, as cited in our group's citation repository for citing the data created and maintained in this repo (NOAA Fisheries Alaska Fisheries Science Center, 2023). Add "note = {Accessed: mm/dd/yyyy}" to append the day this data was accessed.

```
@misc{FOSSAFSCData,
  author = {{NOAA Fisheries Alaska Fisheries Science Center}},
  year = {2023},
  title = {Fisheries One Stop Shop Public Data: RACE Division Bottom Trawl Survey Data Query},
  howpublished = {https://www.fisheries.noaa.gov/foss},
  publisher = {{U.S. Dep. Commer.}},
  copyright = {Public Domain}
}
```

## **9. Data description**

The Resource Assessment and Conservation Engineering Division (RACE) Groundfish Assessment Program (GAP) of the Alaska Fisheries Science Center (AFSC) conducts fisheries-independent bottom trawl surveys to monitor the condition of the demersal fish and crab stocks of Alaska. These data are developed to describe the temporal distribution and abundance of commercially and ecologically important groundfish species, examine the changes in the species composition of the fauna over time and space, and describe the physical environment of the groundfish habitat.

There are no legal restrictions on access to the data. They reside in the public domain and can be freely distributed. Users must read and fully comprehend the metadata prior to use. Data should not be used beyond the limits of the source scale. Acknowledgement of NOAA, as the source from which these data were obtained, in any publications and/or other representations of these data, is suggested. These data are compiled and approved annually after each summer survey season. The data from previous years are unlikely to change substantially once published.

These data are zero-filled (presence and absence) observations from surveys conducted on fishing vessels. These surveys monitor trends in distribution and abundance of groundfish, crab, and bottom-dwelling species in Alaska's marine ecosystems. These data include estimates of catch-per-unit-effort (CPUE) for all identified species for index stations. Some survey data are excluded, such as non-standard stations, surveys completed in earlier years using different/non-standard gear, and special tows and non-standard data collections.

Though not included in the public data, these surveys also collect oceanographic and environmental data, and biological data such as length, weight, stomach contents (to learn more about diet), otoliths (fish ear bones to learn about age), and tissue samples for genetic analysis, all of which can be shared upon special request. Also not included in the public data are estimated biomass (average total weight of all fish and crabs sampled) of crabs and groundfish that support the creation of annual stock assessments.

## *9. Data description*

### **9.1. Data tables**

### **9.2. Data usage**

NA

### **9.3. Data tables**

#### **9.3.1. FOSS\_CATCH**

[There is currently no description for this table.]

Number of rows: 42281918

Number of columns: 12

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

COMMON\_NAME

Taxon Common Name

text

VARCHAR2(255 BYTE)

The common name of the marine organism associated with the 'scientific\_name' and 'species\_code' columns. For a complete species list, review the code books.

COUNT

Taxon Count

count, whole number resolution

NUMBER(38,0)

## *9. Data description*

Total number of individuals caught in haul by taxon, represented in whole numbers used in calculation.

CPUE\_KGKM2

Weight CPUE (kg/km<sup>2</sup>)

kilograms per kilometers squared

NUMBER(38,6)

Catch weight (kilograms) divided by area (squared kilometers) swept by the net.

CPUE\_NOKM2

Number CPUE (no/km<sup>2</sup>)

count per kilometers squared

NUMBER(38,6)

Catch number (in number of organisms) per area (squared kilometers) swept by the net.

HAULJOIN

Haul ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each (vessel, cruise, and haul) combination.

ID\_RANK

Lowest taxonomic rank

text

VARCHAR2(255 BYTE)

Lowest taxonomic rank of a given species entry.

ITIS

ITIS Taxonomic Serial Number

ID code

NUMBER(38,0)

## *9. Data description*

Species code as identified in the Integrated Taxonomic Information System (<https://itis.gov/>).

SCIENTIFIC\_NAME

Taxon Scientific Name

text

VARCHAR2(255 BYTE)

The scientific name of the organism associated with the 'common\_name' and 'species\_code' columns. For a complete taxon list, review the code books.

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

TAXON\_CONFIDENCE

Taxon Confidence Rating

category

VARCHAR2(255 BYTE)

Confidence in the ability of the survey team to correctly identify the taxon to the specified level, based solely on identification skill (e.g., not likelihood of a taxon being caught at that station on a location-by-location basis). Quality codes follow: '**HighModerateLow**

WEIGHT\_KG

Taxon Weight (kg)

## *9. Data description*

kilograms

NUMBER(38,3)

Weight (thousandths of a kilogram) of individuals in a haul by taxon.

WORMS

World Register of Marine Species Taxonomic Serial Number

ID code

NUMBER(38,0)

Species code as identified in the World Register of Marine Species (WoRMS) (<https://www.marinespecies.org/>).

### **9.3.2. FOSS\_CPUE\_PRESONLY**

[There is currently no description for this table.]

Number of rows: 42281918

Number of columns: 37

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_SWEPT\_KM2

Area Swept (km)

kilometers

NUMBER(38,6)

The area the net covered while the net was fishing (kilometers squared), defined as the distance fished times the net width.

BOTTOM\_TEMPERATURE\_C

Bottom Temperature (Degrees Celsius)

degrees Celsius

## *9. Data description*

NUMBER(38,1)

Bottom temperature (tenths of a degree Celsius); NA indicates removed or missing values.

COMMON\_NAME

Taxon Common Name

text

VARCHAR2(255 BYTE)

The common name of the marine organism associated with the 'scientific\_name' and 'species\_code' columns. For a complete species list, review the code books.

COUNT

Taxon Count

count, whole number resolution

NUMBER(38,0)

Total number of individuals caught in haul by taxon, represented in whole numbers used in calculation.

CPUE\_KGKM2

Weight CPUE (kg/km<sup>2</sup>)

kilograms per kilometers squared

NUMBER(38,6)

Catch weight (kilograms) divided by area (squared kilometers) swept by the net.

CPUE\_NOKM2

Number CPUE (no/km<sup>2</sup>)

count per kilometers squared

NUMBER(38,6)

Catch number (in number of organisms) per area (squared kilometers) swept by the net.

CRUISE

Cruise ID

ID code

## *9. Data description*

NUMBER(38,0)

This is a six-digit number identifying the cruise number of the form: YYYY99 (where YYYY = year of the cruise; 99 = 2-digit number and is sequential; 01 denotes the first cruise that vessel made in this year, 02 is the second, etc.).

CRUISEJOIN

Cruise ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each survey, vessel, and year combination.

DATE\_TIME

Date and Time

MM/DD/YYYY HH::MM

DATE

The date (MM/DD/YYYY) and time (HH:MM) of the haul.

DEPTH\_M

Depth (m)

degrees Celsius

NUMBER(38,1)

Bottom depth (tenths of a meter).

DISTANCE\_FISHED\_KM

Distance Fished (km)

degrees Celsius

NUMBER(38,3)

Distance the net fished (thousandths of kilometers).

DURATION\_HR

Tow Duration (decimal hr)

hours

## *9. Data description*

NUMBER(38,1)

This is the elapsed time between start and end of a haul (decimal hours).

HAUL

Haul Number

ID code

NUMBER(38,0)

This number uniquely identifies a sampling event (haul) within a cruise. It is a sequential number, in chronological order of occurrence.

HAULJOIN

Haul ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each (vessel, cruise, and haul) combination.

ID\_RANK

Lowest taxonomic rank

text

VARCHAR2(255 BYTE)

Lowest taxonomic rank of a given species entry.

ITIS

ITIS Taxonomic Serial Number

ID code

NUMBER(38,0)

Species code as identified in the Integrated Taxonomic Information System (<https://itis.gov/>).

LATITUDE\_DD\_END

End Latitude (decimal degrees)

decimal degrees

## *9. Data description*

NUMBER(38,6)

Latitude (one hundred thousandth of a decimal degree) of the end of the haul.

LATITUDE\_DD\_START

Start Latitude (decimal degrees)

decimal degrees

NUMBER(38,6)

Latitude (one hundred thousandth of a decimal degree) of the start of the haul.

LONGITUDE\_DD\_END

End Longitude (decimal degrees)

decimal degrees

NUMBER(38,6)

Longitude (one hundred thousandth of a decimal degree) of the end of the haul.

LONGITUDE\_DD\_START

Start Longitude (decimal degrees)

decimal degrees

NUMBER(38,6)

Longitude (one hundred thousandth of a decimal degree) of the start of the haul.

NET\_HEIGHT\_M

Net Height (m)

meters

NUMBER(38,1)

Measured or estimated distance (meters) between footrope and headrope of the trawl.

NET\_WIDTH\_M

Net Width (m)

meters

NUMBER(38,1)

Measured or estimated distance (meters) between wingtips of the trawl.

## *9. Data description*

### PERFORMANCE

Haul Performance Code

category

NUMBER(38,0)

This denotes what, if any, issues arose during the haul. For more information, review the code books.

### SCIENTIFIC\_NAME

Taxon Scientific Name

text

VARCHAR2(255 BYTE)

The scientific name of the organism associated with the 'common\_name' and 'species\_code' columns. For a complete taxon list, review the code books.

### SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

### SRVY

Survey

text abbreviated

VARCHAR2(255 BYTE)

Abbreviated survey names. The column 'srvy' is associated with the 'survey' and 'survey\_id' columns. Northern Bering Sea (NBS), Southeastern Bering Sea (EBS), Bering Sea Slope (BSS), Gulf of Alaska (GOA), Aleutian Islands (AI).

### STATION

Station ID

ID code

VARCHAR2(255 BYTE)

## *9. Data description*

Alpha-numeric designation for the station established in the design of a survey.

STRATUM

Stratum ID

ID code

NUMBER(10,0)

RACE database statistical area for analyzing data. Strata were designed using bathymetry and other geographic and habitat-related elements. The strata are unique to each survey series. Stratum of value 0 indicates experimental tows.

SURFACE\_TEMPERATURE\_C

Surface Temperature (Degrees Celsius)

degrees Celsius

NUMBER(38,1)

Surface temperature (tenths of a degree Celsius); NA indicates removed or missing values.

SURVEY

Survey Name

text

VARCHAR2(255 BYTE)

Name and description of survey. The column 'survey' is associated with the 'srvy' and 'survey\_id' columns.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

TAXON\_CONFIDENCE

Taxon Confidence Rating

## *9. Data description*

category

VARCHAR2(255 BYTE)

Confidence in the ability of the survey team to correctly identify the taxon to the specified level, based solely on identification skill (e.g., not likelihood of a taxon being caught at that station on a location-by-location basis). Quality codes follow: '**High**': High confidence and consistency. Taxonomy is stable and reliable at this level, and field identification characteristics are well known and reliable. '**Moderate**': Moderate confidence. Taxonomy may be questionable at this level, or field identification characteristics may be variable and difficult to assess consistently. '**Low**': Low confidence. Taxonomy is incompletely known, or reliable field identification characteristics are unknown. Documentation: Species identification confidence in the eastern Bering Sea shelf survey (1982-2008), Species identification confidence in the eastern Bering Sea slope survey (1976-2010), and Species identification confidence in the Gulf of Alaska and Aleutian Islands surveys (1980-2011).

VESSEL\_ID

Vessel ID

ID code

NUMBER(38,0)

ID number of the vessel used to collect data for that haul. The column 'vessel\_id' is associated with the 'vessel\_name' column. Note that it is possible for a vessel to have a new name but the same vessel id number. For a complete list of vessel ID codes, review the code books.

VESSEL\_NAME

Vessel Name

text

VARCHAR2(255 BYTE)

Name of the vessel used to collect data for that haul. The column 'vessel\_name' is associated with the 'vessel\_id' column. Note that it is possible for a vessel to have a new name but the same vessel id number. For a complete list of vessel ID codes, review the code books.

WEIGHT\_KG

Taxon Weight (kg)

kilograms

## *9. Data description*

NUMBER(38,3)

Weight (thousandths of a kilogram) of individuals in a haul by taxon.

WORMS

World Register of Marine Species Taxonomic Serial Number

ID code

NUMBER(38,0)

Species code as identified in the World Register of Marine Species (WoRMS) (<https://www.marinespecies.org/>).

YEAR

Year

year

NUMBER(10,0)

Year the survey was conducted in.

### **9.3.3. FOSS\_HAUL**

[There is currently no description for this table.]

Number of rows: 32510

Number of columns: 27

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AREA\_SWEPT\_KM2

Area Swept (km)

kilometers

NUMBER(38,6)

## *9. Data description*

The area the net covered while the net was fishing (kilometers squared), defined as the distance fished times the net width.

BOTTOM\_TEMPERATURE\_C

Bottom Temperature (Degrees Celsius)

degrees Celsius

NUMBER(38,1)

Bottom temperature (tenths of a degree Celsius); NA indicates removed or missing values.

CRUISE

Cruise ID

ID code

NUMBER(38,0)

This is a six-digit number identifying the cruise number of the form: YYYY99 (where YYYY = year of the cruise; 99 = 2-digit number and is sequential; 01 denotes the first cruise that vessel made in this year, 02 is the second, etc.).

CRUISEJOIN

Cruise ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each survey, vessel, and year combination.

DATE\_TIME

Date and Time

MM/DD/YYYY HH::MM

DATE

The date (MM/DD/YYYY) and time (HH:MM) of the haul.

DEPTH\_M

Depth (m)

degrees Celsius

## *9. Data description*

NUMBER(38,1)

Bottom depth (tenths of a meter).

DISTANCE\_FISHED\_KM

Distance Fished (km)

degrees Celsius

NUMBER(38,3)

Distance the net fished (thousandths of kilometers).

DURATION\_HR

Tow Duration (decimal hr)

hours

NUMBER(38,1)

This is the elapsed time between start and end of a haul (decimal hours).

HAUL

Haul Number

ID code

NUMBER(38,0)

This number uniquely identifies a sampling event (haul) within a cruise. It is a sequential number, in chronological order of occurrence.

HAULJOIN

Haul ID

ID code

NUMBER(38,0)

This is a unique numeric identifier assigned to each (vessel, cruise, and haul) combination.

LATITUDE\_DD\_END

End Latitude (decimal degrees)

decimal degrees

NUMBER(38,6)

## *9. Data description*

Latitude (one hundred thousandth of a decimal degree) of the end of the haul.

LATITUDE\_DD\_START

Start Latitude (decimal degrees)

decimal degrees

NUMBER(38,6)

Latitude (one hundred thousandth of a decimal degree) of the start of the haul.

LONGITUDE\_DD\_END

End Longitude (decimal degrees)

decimal degrees

NUMBER(38,6)

Longitude (one hundred thousandth of a decimal degree) of the end of the haul.

LONGITUDE\_DD\_START

Start Longitude (decimal degrees)

decimal degrees

NUMBER(38,6)

Longitude (one hundred thousandth of a decimal degree) of the start of the haul.

NET\_HEIGHT\_M

Net Height (m)

meters

NUMBER(38,1)

Measured or estimated distance (meters) between footrope and headrope of the trawl.

NET\_WIDTH\_M

Net Width (m)

meters

NUMBER(38,1)

Measured or estimated distance (meters) between wingtips of the trawl.

PERFORMANCE

## *9. Data description*

Haul Performance Code

category

NUMBER(38,0)

This denotes what, if any, issues arose during the haul. For more information, review the code books.

SRVY

Survey

text abbreviated

VARCHAR2(255 BYTE)

Abbreviated survey names. The column 'srvy' is associated with the 'survey' and 'survey\_id' columns. Northern Bering Sea (NBS), Southeastern Bering Sea (EBS), Bering Sea Slope (BSS), Gulf of Alaska (GOA), Aleutian Islands (AI).

STATION

Station ID

ID code

VARCHAR2(255 BYTE)

Alpha-numeric designation for the station established in the design of a survey.

STRATUM

Stratum ID

ID code

NUMBER(10,0)

RACE database statistical area for analyzing data. Strata were designed using bathymetry and other geographic and habitat-related elements. The strata are unique to each survey series. Stratum of value 0 indicates experimental tows.

SURFACE\_TEMPERATURE\_C

Surface Temperature (Degrees Celsius)

degrees Celsius

NUMBER(38,1)

## *9. Data description*

Surface temperature (tenths of a degree Celsius); NA indicates removed or missing values.

SURVEY

Survey Name

text

VARCHAR2(255 BYTE)

Name and description of survey. The column 'survey' is associated with the 'srvy' and 'survey\_id' columns.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

This number uniquely identifies a survey. Name and description of survey. The column 'survey\_id' is associated with the 'srvy' and 'survey' columns. For a complete list of surveys, review the code books.

SURVEY\_NAME

NA

NA

NA

NA

VESSEL\_ID

Vessel ID

ID code

NUMBER(38,0)

ID number of the vessel used to collect data for that haul. The column 'vessel\_id' is associated with the 'vessel\_name' column. Note that it is possible for a vessel to have a new name but the same vessel id number. For a complete list of vessel ID codes, review the code books.

VESSEL\_NAME

Vessel Name

## *9. Data description*

text

VARCHAR2(255 BYTE)

Name of the vessel used to collect data for that haul. The column 'vessel\_name' is associated with the 'vessel\_id' column. Note that it is possible for a vessel to have a new name but the same vessel id number. For a complete list of vessel ID codes, review the code books.

YEAR

Year

year

NUMBER(10,0)

Year the survey was conducted in.

### **9.3.4. FOSS\_TAXON\_GROUP**

[There is currently no description for this table.]

Number of rows: 37606

Number of columns: 3

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

CLASSIFICATION

Taxonomic classification rank group

category

VARCHAR2(255 BYTE)

Phylogenetic classification group rank for a given species.

RANK\_ID

NA

*9. Data description*

NA

NA

NA

SPECIES\_CODE

Taxon Code

ID code

NUMBER(38,0)

The species code of the organism associated with the 'common\_name' and 'scientific\_name' columns. For a complete species list, review the code books.

# 10. Using the API

## 10.1. Select and filter

Select, filter, and package this and other NOAA Fisheries data from the Fisheries One Stop Shop (FOSS) platform. A user guide for the FOSS platform can be found here. To begin a report, select options from the boxes what you need data for.

For a given box, select one or a few options from the “options box” (list on the left) to query by highlighting them. To select multiple options, hold down the CTRL key while clicking on the options of interest, or click and drag down the list. Once the options you wish to be included in your query are highlighted, click the right-pointing arrow (>) to move them into the “selection box” (list on the right). If you accidentally select an option that you do not want to query, simply select the unwanted option from the selection box and click the left-pointing arrow (<).

If you wish to select all options from the options box and send them to the selection box, simply click the double right-pointing arrow (>>). If you want to unselect all options from the selection box, use the double left-pointing arrow (<<) or the reset icon.

To find a specific species or group more quickly you can use the `Search Species` option to quickly narrow the options. Search for parts of species common names in the `Search Species` box by entering a term and clicking the `search` button. The platform will return a shorter list in the `Species` options box of only species that contain a match to that search term.

Use the `Reset All Parameters` button to reset all parameters for entire form.

Filter options:

- **Survey:** Each survey has different in design, time series, and history. More information on each survey and their designs can be found in our annual data reports.
- **Year:** Surveys are not conducted in all years, so only data from the years for which the survey was conducted will be returned.

## 10. Using the API

The screenshot shows the AFSC Groundfish and Crab Assessment Program Bottom Trawl Surveys data interface. At the top, the NOAA Fisheries logo is visible. The main title "AFSC Groundfish and Crab Assessment Program" and "Bottom Trawl Surveys" is displayed above a banner image of a fish swimming in water. Below the title, a message states "Survey data also available through API". On the left, a sidebar menu includes links to FUS Report, Landings, Foreign Trade, Top US Ports, Processed Products, Per Capita Consumption, Supply, USCG Vessels, AFSC GAP Survey (which is selected), AFSC GAP Metadata, Partners, Metadata and Caveats, Frequently Asked Questions, Quick Start Guide, and Comments. The main content area has a "Data Caveats" section with a note about survey data being presence-only. A "Parameters" section contains dropdown menus for Survey (with options like Aleutian Islands Bottom Trawl Survey, Eastern Bering Sea Slope Bottom Trawl Survey, etc.), Year (from 2014 to 2022), and Species (including Abalone Jingle, Abietinaria, Abyssal Crangon, Acanthascus, Acantholithodes, Achanax johnsoni, Acoel Turbellarian, Acteocina). There is also a "Search Species" input field and a "RUN REPORT" button at the bottom.

Figure 10.1.: AFSC Groundfish and Crab Assessment Program Bottom Trawl Survey data interface on the Fisheries One Stop Shop platform.

## 10. Using the API

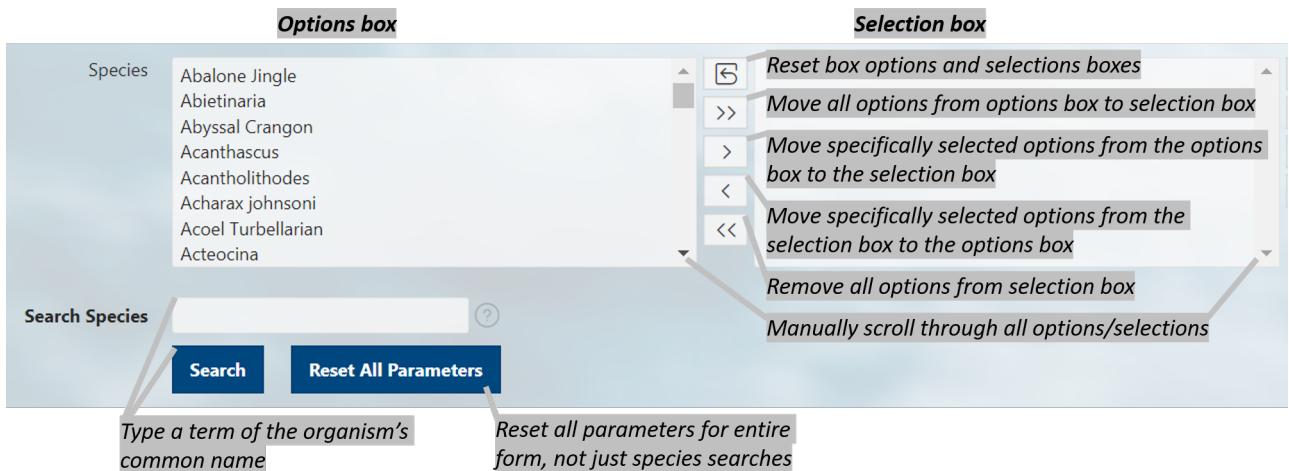


Figure 10.2.: Diagram of selection and search tools available on the FOSS platoform.

- **Species:** Common name of all species ever encountered in the survey. Find more information about these species in our survey code books.

In this example, we'll select for 2022 eastern Bering Sea Pacific cod data. Here, we used the **Search Species** box to search for species with the term "cod" in their common names and selected "Pacific cod" from that shortened list.

### 10.2. Select data format

Select from the below radio list of pre-designed output tables. Once you run the report, the user can further specify filter data and select columns of interest. The tables below will only include data from the selections made in the previous step.

- **All Data Fields: Presence and Absence (zero-filled):** The most complete version of the data, including species, catch, haul, and environmental data. This data will include catch data for where species were caught and zeros for where the species were not caught. This is important for calculating catch-per-unit-effort data, preparing distribution plots (e.g., using the akgfmaps R package), and many statistical analyses.
- **All Data Fields: Presence-only (non-zero):** The second most complete version of the data, including species, catch, haul, and environmental data. However, this data only includes catch data for where species were caught and does

## 10. Using the API

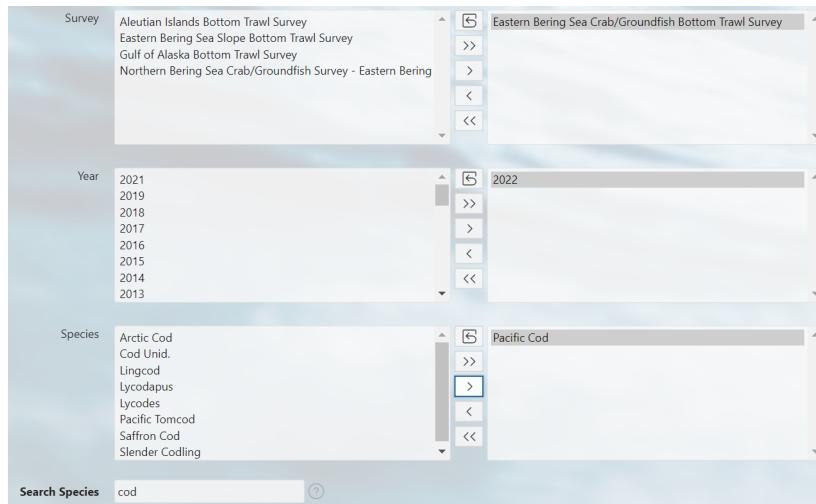


Figure 10.3.: Diagram of selection and search tools available on the FOSS platoform.

not include zeros for where the species were not caught. This will return smaller, more focused data and can be useful for quickly assessing how many species were caught or how many stations species were caught at.

- Catch data: Presence and Absence (zero-filled): This data set is similar to All Data Fields: Presence and Absence (zero-filled), but only includes catch and species data columns.
- Catch data: Presence-only (non-zero): This data set is similar to All Data Fields: Presence-only (non-zero), but only includes catch and species data columns.
- Haul Data: This data set only includes haul and environmental data collected from the survey. This data will only include one observation per haul event/station.

In this example, we'll select All Data Fields: Presence and Absence (zero-filled).

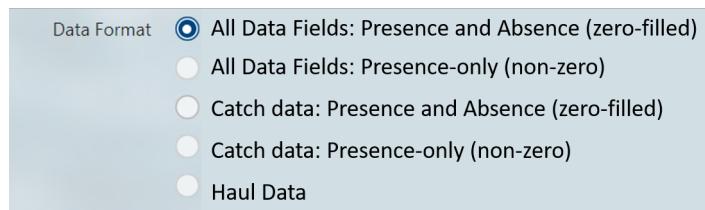


Figure 10.4.: Diagram of the pre-set data format options.

## 10. Using the API

### 10.3. Run report

Click the **RUN REPORT** button. Below the select and filter area, the results of your query will appear below the page in the format you selected. To change the format, make a different selection and run the report again. Further modifications to your results can be made by clicking on the **Actions** button above your data. Here you can download your data, select columns included in your results, and apply a variety of filters and mathematical tools.

Q		Go	Rows 50	Actions		
Year	Srvy	Survey	Survey Id	Cruise	Help	Select Columns
2022	EBS	Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey	9.8E+001	2.02201E+005	1.0E-005	Filter
2022	EBS	Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey	9.8E+001	2.02201E+005	1.0E-005	Rows Per Page
2022	EBS	Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey	9.8E+001	2.02201E+005	1.0E-005	Format
2022	EBS	Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey	9.8E+001	2.02201E+005	1.0E-005	Flashback
2022	EBS	Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey	9.8E+001	2.02201E+005	1.0E-005	Reset
2022	EBS	Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey	9.8E+001	2.02201E+005	1.1E-005	Help
2022	EBS	Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey	9.8E+001	2.02201E+005	1.1E-005	Download

Figure 10.5.: Example data returned from running the report.

# 11. Access API data using R

An application programming interface (API) is a way for two or more computer programs to communicate with each other.

More information about how to amend API links can be found [here](#). Useful introductions to using APIs in R can be found [here](#).

## 11.1. Ex. 1: Load the first 25 rows (default) of data

```
# install.packages(c("httr", "jsonlite"))
library(httr)
library(jsonlite)
library(dplyr)

# link to the API
api_link <- "https://apps-st.fisheries.noaa.gov/ods/foss/afsc_groundfish_survey/"

res <- httr::GET(url = api_link)
# res # Test connection
data <- jsonlite::fromJSON(base::rawToChar(res$content))
# names(data)
tibble::as_tibble(data$items) %>%
  dplyr::mutate_if(is.character, type.convert, as.is = TRUE) %>%
  dplyr::mutate(across(where(is.numeric), round, 3)) %>%
  head(3) %>%
flextable::flextable() %>%
  flextable::theme_zebra() %>%
  flextable::colformat_num(x = ., j = c("year", "cruise", "species_code", "tsn", "ak_surv
```

## 11. Access API data using R

yearsrvy	survey	survey_id	cruise	haul	stratumstation	vessel_name	vessel
2002AI	Aleutian Islands Bottom Trawl Survey	52	200201	6	722307-63	Vesteraale	
2002AI	Aleutian Islands Bottom Trawl Survey	52	200201	6	722307-63	Vesteraalen	
2002AI	Aleutian Islands Bottom Trawl Survey	52	200201	6	722307-63	Vesteraale	

### 11.2. Ex. 2: Load the first 10000 rows of data

```
# Not run because too big:  
res <- httr::GET(url = paste0(api_link, "?offset=0&limit=10000"))  
data <- jsonlite::fromJSON(base::rawToChar(res$content))  
print(paste0("rows: ", dim(data$items)[1], "; cols: ", dim(data$items)[2]))
```

```
[1] "rows: 10000; cols: 36"
```

### 11.3. Ex. 3: Filter by Year

Show all the data greater than the year 2020.

## 11. Access API data using R

```
res <- httr::GET(url = paste0(api_link, '?q={"year":{$gt":2020}}'))  
data <- jsonlite::fromJSON(base::rawToChar(res$content))  
  
as_tibble(data$items) %>%  
  mutate_if(is.character, type.convert, as.is = TRUE) %>%  
  head(3) %>%  
  dplyr::mutate(across(where(is.numeric), round, 3)) %>%  
  dplyr::select(year, srvy, stratum, species_code, cpue_kgkm2) %>%  
  flextable::flextable() %>%  
  flextable::theme_zebra() %>%  
  flextable::colformat_num(x = ., j = c("year", "species_code"), big.mark = "")
```

Table 11.2.: Ex. 3: Filter by Year.

year	srvy	stratum	species_code	cpue_kgkm2
2022	AI	722	10261	673.326
2022	AI	793	80540	0.361
2022	AI	722	21347	758.091

### 11.4. Ex. 4: Filter by species name

Show all the data where the product name contains pollock Please note that here the word pollock is case sensitive.

The notation for finding a string is to use % around it. Since % is a reserved character in a URL, you have to replace % with %25.

```
res <- httr::GET(  
  url = paste0(api_link, '?q={"common_name":{"$like":"%25pollock%25"}}'))  
data <- jsonlite::fromJSON(base::rawToChar(res$content))  
  
as_tibble(data$items) %>%  
  mutate_if(is.character, type.convert, as.is = TRUE) %>%  
  head(3) %>%  
  dplyr::mutate(across(where(is.numeric), round, 3)) %>%
```

## 11. Access API data using R

```
dplyr::select(year, srvy, stratum, species_code, cpue_kgkm2) %>%
flextable::flextable() %>%
  flextable::theme_zebra() %>%
  flextable::colformat_num(x = ., j = c("year", "species_code"), big.mark = "")
```

Table 11.3.: Ex. 4: Filter by species name.

yearsrvy	stratum	species_code	cpue_kgkm2
2002AI	721	21740	0.640
2002AI	722	21740	775.322
2002AI	722	21740	10,685.806

### 11.5. Ex. 5: Combination of year and name filters

Show all the data where years > 2020 and the product name contains pollock

```
res <- httr::GET(
  url = paste0(api_link,
               '?q={"year":{"$gt":2020}, "common_name":{"$like": "%pollock%"} }'))
data <- jsonlite::fromJSON(base::rawToChar(res$content))

as_tibble(data$items) %>%
  mutate_if(is.character, type.convert, as.is = TRUE) %>%
  head(3) %>%
  dplyr::mutate(across(where(is.numeric), round, 3)) %>%
  dplyr::select(year, srvy, stratum, species_code, cpue_kgkm2) %>%
  flextable::flextable() %>%
    flextable::theme_zebra() %>%
    flextable::colformat_num(x = ., j = c("year", "species_code"), big.mark = "")
```

## 11. Access API data using R

Table 11.4.: Ex. 5: Combination of year and name filters.

<b>yearsrvy</b>	<b>stratum</b>	<b>species_- code</b>	<b>cpue_- kgkm2</b>
2022AI	722	2174022	754.334
2022AI	793	21740	7,853.632
2022AI	721	21740	7,235.010

## 11.6. Ex. 6: Combination of year, srvy, stratum

Show all the data where year = 1989, srvy = "EBS", and stratum is not equal to 81

```
res <- httr::GET(
  url = paste0(api_link, '?q={"year":1989,"srvy":"EBS","stratum":{"$ne":"81"}})')
data <- jsonlite::fromJSON(base::rawToChar(res$content))

as_tibble(data$items) %>%
  mutate_if(is.character, type.convert, as.is = TRUE) %>%
  head(3) %>%
  dplyr::mutate(across(where(is.numeric), round, 3)) %>%
  dplyr::select(year, srvy, stratum, species_code, cpue_kgkm2) %>%
  flextable::flextable() %>%
  flextable::theme_zebra() %>%
  flextable::colformat_num(x = ., j = c("year", "species_code"), big.mark = "")
```

Table 11.5.: Ex. 6: Combination of year, srvy, stratum.

<b>yearsrvy</b>	<b>stratum</b>	<b>species_- code</b>	<b>cpue_- kgkm2</b>
1989EBS	10	40500	9.620
1989EBS	10	68578	9.620
1989EBS	10	21313	18.179

## 11.7. Ex. 7: Visualize CPUE data in distribution map

Pacific cod catch-per-unit-effort estimates for NBS in 2021 and map constructed using akfmaps.

```
# res <- httr::GET(
#   url = paste0(api_link, "?offset=0&limit=10000"),
#   query = list(year = 2021, srvy = "EBS", species_code = 30060))
res <- httr::GET(
  url = paste0(api_link, '?q={"year":2021,"srvy":"NBS","species_code":21720}'))
data_catch <- jsonlite::fromJSON(base::rawToChar(res$content))$items %>%
  dplyr::select(stratum, station, cpue_kgkm2)

# zero-fill data (imperfectly, but effective for this example)
res <- httr::GET(
  url = paste0(api_link, '?q={"year":2021,"srvy":"NBS"}offset=0&limit=10000'))
data_haul <- jsonlite::fromJSON(base::rawToChar(res$content))$items %>%
  dplyr::select(stratum, station, latitude_dd, longitude_dd) %>%
  dplyr::mutate(across(where(is.numeric), round, 3)) %>%
  dplyr::distinct()

data <- dplyr::left_join(data_haul, data_catch) %>%
  dplyr::mutate(cpue_kgkm2 = ifelse(is.na(cpue_kgkm2), 0, cpue_kgkm2),
    dplyr::across(dplyr::everything(), as.numeric))

flextable::flextable(data[1:3,]) %>%
  flextable::theme_zebra()
```

Table 11.6.: Ex. 7: Visualize CPUE data in distribution map.

stratum	station	latitude_dd	longitude_dd	cpue_kgkm2
71		63.70028	-171.0225	1.183039
81		61.68600	-173.07761	3,256.716473
81		61.34965	-172.22516	0.958261

## 11. Access API data using R

```
# devtools::install_github("afsc-gap-products/akgfmmaps", build_vignettes = TRUE)
library(akgfmmaps)

figure <- akgfmmaps::make_idw_map(
  CPUE_KGHA = data$cpue_kgkm2, # calculates the same, regardless of units.
  LATITUDE = data$latitude_dd,
  LONGITUDE = data$longitude_dd,
  region = "bs.north", # Predefined EBS area
  set.breaks = "jenks", # Gets Jenks breaks from classint::classIntervals()
  in.crs = "+proj=longlat", # Set input coordinate reference system
  out.crs = "EPSG:3338", # Set output coordinate reference system
  grid.cell = c(20000, 20000), # 20x20km grid
  key.title = "Pacific Ocean perch") # Include in the legend title

[inverse distance weighted interpolation]
[inverse distance weighted interpolation]

figure$plot +
  ggplot2::guides(fill=guide_legend(title = "Pacific cod\nCPUE (kg/km2)"))
```

## 11. Access API data using R

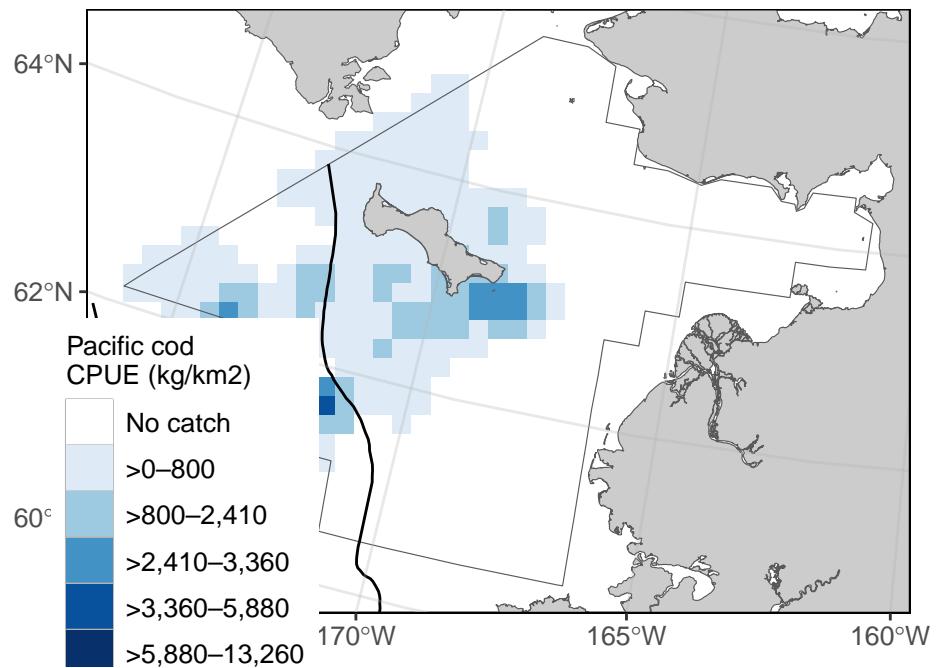


Figure 11.1.: Ex. 7: Visualize CPUE data in distribution map.

# 12. Access API data using Python

## 12.0.1. {afscgap} Library Installation

author: Sam Pottinger (sam.pottinger@berkeley.edu; GitHub::sampottinger)  
date: May 13, 2023

The third-party afscgap Python package interfaces with FOSS to access AFSC GAP data. It can be installed via pip:

```
#The reticulate package provides a comprehensive set of tools for interoperability between
library(reticulate)
```

```
pip install afscgap
pip install git+https://github.com/SchmidtDSE/afscgap.git@main
```

For more information on installation and deployment, see the library documentation.

## 12.0.2. Basic query

This first example queries for Pacific glass shrimp (*Pasiphaea pacifica*) in the Gulf of Alaska in 2021. The library will automatically generate HTTP queries, converting from Python types to ORDS query syntax.

```
import afscgap

query = afscgap.Query()
query.filter_year(eq=2021)
query.filter_srvy(eq='GOA')
query.filter_scientific_name(eq='Pasiphaea pacifica')

results = query.execute()
```

## 12. Access API data using Python

The `results` variable in this example is an iterator that will automatically perform pagination behind the scenes.

### 12.0.3. Iterating with a for loop

The easiest way to interact with results is a simple for loop. This next example determines the frequency of different catch per unit effort where Pacific glass shrimp were reported:

```
import afscgap

# Mapping from CPUE to count
count_by_cpue = {}

# Build query
query = afscgap.Query()
query.filter_year(eq=2021)
query.filter_srvy(eq='GOA')
query.filter_scientific_name(eq='Pasiphaea pacifica')
results = query.execute()

# Iterate through results and count
for record in results:
    cpue = record.get_cpue_weight(units='kg/ha')
    cpue_rounded = round(cpue)
    count = count_by_cpue.get(cpue_rounded, 0) + 1
    count_by_cpue[cpue_rounded] = count

# Print the result
print(count_by_cpue)
```

Note that, in this example, only records with Pacific glass shrimp are included (“presence-only” data). See zero catch inference below. In other words, it reports on CPUE only for hauls in which Pacific glass shrimp were recorded, excluding some hauls like those in which Pacific glass shrimp were not found at all.

## 12. Access API data using Python

### 12.0.4. Iterating with functional programming

A for loop is not the only option for iterating through results. List comprehensions and other functional programming methods can be used as well.

```
import statistics

import afscgap

# Build query
query = afscgap.Query()
query.filter_year(eq=2021)
query.filter_srvy(eq='GOA')
query.filter_scientific_name(eq='Pasiphaea pacifica')
results = query.execute()

# Get temperatures in Celsius
temperatures = [record.get_bottom_temperature(units='c') for record in results]

# Take the median
print(statistics.median(temperatures))
```

This example reports the median temperature in Celcius for when Pacific glass shrimp was reported.

### 12.0.5. Load into Pandas

The results from the afscgap package are serializable and can be loaded into other tools like Pandas. This example loads Pacific glass shrimp from 2021 Gulf of Alaska into a data frame.

```
import pandas

import afscgap

query = afscgap.Query()
query.filter_year(eq=2021)
query.filter_srvy(eq='GOA')
query.filter_scientific_name(eq='Pasiphaea pacifica')
```

## 12. Access API data using Python

```
results = query.execute()  
  
pandas.DataFrame(results.to_dicts())
```

Specifically, `to_dicts` provides an iterator over a dictionary form of the data that can be read into tools like Pandas.

### 12.0.6. Advanced filtering

Queries so far have focused on filters requiring equality but range queries can be built as well.

```
import afscgap  
  
# Build query  
query = afscgap.Query()  
query.filter_year(min_val=2015, max_val=2019)    # Note min/max_val  
query.filter_srvy(eq='GOA')  
query.filter_scientific_name(eq='Pasiphaea pacifica')  
results = query.execute()  
  
# Sum weight  
weights = map(lambda x: x.get_weight(units='kg'), results)  
total_weight = sum(weights)  
print(total_weight)
```

This example queries for Pacific glass shrimp data between 2015 and 2019, summing the total weight caught. Note that most users will likely take advantage of built-in Python to ORDS query generation which dictates how the library communicates with the API service. However, users can provide raw ORDS queries as well using manual filtering.

### 12.0.7. Zero-catch inference

Until this point, these examples use presence-only data. However, the `afscgap` package can infer negative or “zero catch” records as well.

## 12. Access API data using Python

```
import afscgap

# Mapping from CPUE to count
count_by_cpue = {}

# Build query
query = afscgap.Query()
query.filter_year(eq=2021)
query.filter_srvy(eq='GOA')
query.filter_scientific_name(eq='Pasiphaea pacifica')
query.set_presence_only(False) # Added to earlier example
results = query.execute()

# Iterate through results and count
for record in results:
    cpue = record.get_cpue_weight(units='kg/ha')
    cpue_rounded = round(cpue)
    count = count_by_cpue.get(cpue_rounded, 0) + 1
    count_by_cpue[cpue_rounded] = count

# Print the result
print(count_by_cpue)
```

This example revisits the earlier snippet for CPUE counts but `set_presence_only(False)` directs the library to look at additional data on hauls, determining which hauls did not have Pacific glass shrimp. This lets the library return records for hauls in which Pacific glass shrimp were not found. This can be seen in differences in counts reported:

Rounded CPUE	Count with <code>set_presence_only(True)</code>	Count with <code>set_presence_only(False)</code>
0 kg/ha	44	521
1 kg/ha	7	7
2 kg/ha	1	1

Put simply, while the earlier example showed CPUE counts for hauls in which Pacific glass shrimp were seen, this revised example reports for all hauls in the Gulf of Alaska in 2021.

## *12. Access API data using Python*

### **12.0.8. More information**

Please see the API documentation for the Python library for additional details.

## 13. Access data using R (AFSC only)

If the user has access to the AFSC Oracle database, the user can use SQL developer to view and pull the FOSS public data directly from the RACEBASE\_FOSS Oracle schema.

### 13.0.1. Connect to Oracle from R

Many users will want to access the data from Oracle using R. The user will need to install the RODBC R package and ask OFIS (IT) connect R to Oracle. Then, use the following code in R to establish a connection from R to Oracle:

Here, the user can write in their username and password directly into the RODBC connect function. Never save usernames or passwords in scripts that may be intentionally or unintentionally shared with others. If no username and password is entered in the function, pop-ups will appear on the screen asking for the username and password.

```
#' Define RODBC connection to ORACLE
#'
#' @param schema default = 'AFSC'.
#'
#' @return oracle channel connection
#' @export
#'
#' @examples
#' # Not run
#'
#' # channel <- oracle_connect()
oracle_connect <- function(
  schema='AFSC',
  username = NULL,
  passowrd = NULL){(echo=FALSE)

  library("RODBC")
  library("getPass")
  if (is.null(username)) {
```

### 13. Access data using R (AFSC only)

```
username <- getPass(msg = "Enter your ORACLE Username: ")
}
if (is.null(password)) {
  password <- getPass(msg = "Enter your ORACLE Password: ")
}
channel <- RODBC::odbcConnect(
  paste(schema),
  paste(username),
  paste(password),
  believeNRows=FALSE)
return(channel)
}

channel <- oracle_connect()
```

#### 13.0.2. Ex. 1: Join data

To join these tables in Oracle, you may use a variant of the following code:

#### 13.0.3. Ex. 2: Subset data

Once connected, pull and save (if needed) the tables into the R environment.

To pull a small subset of the data (especially since files like RACEBASE\_FOSS.FOSS\_-CPUE\_ZEROILLED are so big), use a variation of the following code. Here, we are pulling EBS Pacific cod from 2010 - 2021:

```
# Pull data
a <- RODBC::sqlQuery(
  channel = channel,
  query =
  "SELECT * FROM GAP_PRODUCTS.FOSS_CATCH cc
  JOIN GAP_PRODUCTS.FOSS_HAUL hh
  ON cc.HAULJOIN = hh.HAULJOIN
  WHERE SRVY = 'EBS'
  AND COMMON_NAME = 'Pacific cod'
  AND YEAR >= 2010
  AND YEAR < 2021")
```

*13. Access data using R (AFSC only)*

```
# Save table to local directory  
write.csv(x = a, file = "RACEBASE_FOSS-FOSS_CPUE_ZEROFILED-ebs_pcod_2010-2020.csv")
```

## **14. Other products**

### **14.1. Annual Publications**

#### **14.1.1. Annual Data Report Tech Memos**

The Alaska Fisheries Science Center uses the NOAA Technical Memorandum series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible. Documents within this series reflect sound professional work and may be referenced in the formal scientific and technical literature. These are available online and the latest publications for each survey are listed below.

**POCs:** Bering Sea: Emily Markowitz, Liz Dawson, Chris Anderson Gulf of Alaska and Aleutian Islands: margaret Siple, Alex Dowlin

#### **14.1.2. Community Highlights**

Each year, the Groundfish Assessment Program compiles its survey findings for communities around Alaska (Markowitz et al., 2022). *Please note: This document is for informational purposes only and does not necessarily represent the views or official position of the Department of Commerce, the National Oceanic and Atmospheric Administration, or the National Marine Fisheries Service. Not to be cited without permission from the authors.*

**POCs:** Bering Sea: Emily Markowitz, Liz Dawson, Chris Anderson

#### **14.1.3. North Pacific Groundfish Plan Team**

Each year, the survey teams present their findings to the North Pacific Groundfish Plan Team. Find those presentations, recordings, and attachments on the North Pacific Fishery Management Council website. ### Research Briefs

## *14. Other products*

Each year, the Groundfish Assessment Program produces research briefs, or research plans, of their upcoming surveys and research.

### **14.1.4. Code Books**

The Species and Gear Code book is a listing of codes used for fish and invertebrates identified in RACE Division surveys.

**POCs:** Nancy Roberson

### **14.1.5. Survey Protocols**

Groundfish bottom trawl survey protocols are documented in NOAA protocols for groundfish bottom trawl surveys of the nation's fishery resources.

**POCs:** Nancy Roberson

## **14.2. NOAA Fisheries Distribution Mapping and Analysis Portal (DisMAP)**

Data from this and other fisheries-independent surveys are used in the NOAA Fisheries Distribution Mapping and Analysis Portal (DisMAP), which provides easy access to information to track and understand distributions of marine species in U.S. Marine Ecosystems.

**POCs:** NOAA Fisheries Office of Science and Technology

## **14.3. R Packages**

- akgfmaps R package: bottom trawl survey maps layers and plotting examples.  
**POC:** Sean Rohan
- coldpool R package: cold pool area and temperature data products for the Bering Sea. **POC:** Sean Rohan
- akfishcondition R package: groundfish morphometric condition indicators for fish in the Bering Sea, Aleutian Islands, and Gulf of Alaska. **POC:** Sean Rohan

#### *14. Other products*

- gapindex R package: calculation of Design-Based Indices of Abundance and Composition for AFSC GAP Bottom Trawl Surveys. **POC:** Zack Oyafuso

### **14.4. Outreach**

Maps of near real-time temperatures and survey progress from each of our bottom trawl surveys are posted each business day of the survey. Water temperature affects many species' spawning times, access to food, growth rates, and overall range. Collecting temperature data helps better understand species' habitats and the larger ecosystem. **POCs:** Bering Sea: Emily Markowitz, Liz Dawson, Chris Anderson

### **14.5. Collaborators**

Our data are used in many annual publications, including but not limited to the list below:

- Alaska Stock Assessments
- North Pacific Groundfish Stock Assessment and Fishery Evaluation Reports
- Groundfish Economic Status Reports for the Gulf of Alaska and Bering Sea and Aleutian Islands
- Alaska Marine Ecosystem Status Report Database
- Southeast Alaska Coastal Monitoring Survey Reports
- Alaska Fisheries Life History Database
- Essential Fish Habitat Research Plan in Alaska

## **Part VI.**

### **Contact us**

### *Suggestions and comments*

**General questions and more specific data requests** can be sent to afsc.gap. metadata@noaa.gov or submitted as an issue on our GitHub Organization. The version of this data used for stock assessments can be found through the Alaska Fisheries Information Network (AKFIN). For questions about the eastern Bering Sea surveys, contact Duane Stevenson (Duane.Stevenson@noaa.gov). For questions about the Gulf of Alaska or Aleutian Islands surveys, contact Ned Laman (Ned.Laman@noaa.gov). For questions specifically about crab data in any region, contact Mike Litzow (Mike.Litzow@noaa.gov), the Shellfish Assessment Program lead.

For questions, comments, and concerns specifically about the Fisheries One Stop Shop (FOSS) platform, please contact us using the Comments page on the FOSS web-page.

Alaska Fisheries Science Center (AFSC)  
National Oceanic and Atmospheric Administration (NOAA)  
Resource Assessment and Conservation Engineering Division (RACE)  
Groundfish Assessment Program (GAP)  
7600 Sand Point Way, N.E. bldg. 4  
Seattle, WA 98115 USA

## **Suggestions and comments**

If the data or metadata can be improved, please create a pull request, submit an issue to the GitHub organization or submit an issue to the code's repository.

## **15. Production Run Notes**

## 16. R Version Metadata

```
R version 4.3.0 (2023-04-21 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default
```

```
locale:
[1] LC_COLLATE=English_United States.utf8
[2] LC_CTYPE=English_United States.utf8
[3] LC_MONETARY=English_United States.utf8
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.utf8

time zone: America/Los_Angeles
tzcode source: internal

attached base packages:
[1] stats      graphics   grDevices utils      datasets   methods    base

loaded via a namespace (and not attached):
[1] compiler_4.3.0    fastmap_1.1.1    cli_3.6.1      tools_4.3.0
[5] htmltools_0.5.5   rstudioapi_0.15.0  yaml_2.3.7    rmarkdown_2.23
[9] knitr_1.43       jsonlite_1.8.7   xfun_0.39     digest_0.6.33
[13] rlang_1.1.1      evaluate_0.21
```

### 16.0.1. NOAA README

This repository is a scientific product and is not official communication of the National Oceanic and Atmospheric Administration, or the United States Department of Commerce. All NOAA GitHub project code is provided on an ‘as is’ basis and the user assumes responsibility for its use. Any claims against the Department of Commerce or

## *16. R Version Metadata*

Department of Commerce bureaus stemming from the use of this GitHub project will be governed by all applicable Federal law. Any reference to specific commercial products, processes, or services by service mark, trademark, manufacturer, or otherwise, does not constitute or imply their endorsement, recommendation or favoring by the Department of Commerce. The Department of Commerce seal and logo, or the seal and logo of a DOC bureau, shall not be used in any manner to imply endorsement of any commercial product or activity by DOC or the United States Government.

### **16.0.2. NOAA License**

Software code created by U.S. Government employees is not subject to copyright in the United States (17 U.S.C. §105). The United States/Department of Commerce reserve all rights to seek and obtain copyright protection in countries other than the United States for Software authored in its entirety by the Department of Commerce. To this end, the Department of Commerce hereby grants to Recipient a royalty-free, nonexclusive license to use, copy, and create derivative works of the Software outside of the United States.

## 17. Data constraints

### 17.1. Cite this data

Use the below bibtext citations, as cited in our group's citation repository for citing the data created and maintained in this repo. Add "note = {Accessed: mm/dd/yyyy}" to append the day this data was accessed. Included here are AFSC RACE Groundfish and Shellfish Assessment Program's:

- Design-Based Production Data internal.
- AFSC RACE Groundfish Data for AKFIN.
- Public Data hosted on the Fisheries One Stop Shop (FOSS) Data Platform.

```
\n@misc{GAPPProducts,\n  author = {{NOAA Fisheries Alaska Fisheries Science Center, Goundfis}}
```

## 18. Access Constraints

There are no legal restrictions on access to the data. They reside in public domain and can be freely distributed.

**User Constraints:** Users must read and fully comprehend the metadata prior to use. Data should not be used beyond the limits of the source scale. Acknowledgement of AFSC Groundfish Assessment Program, as the source from which these data were obtained, in any publications and/or other representations of these data, is suggested.

## **19. Acknowledgments**

## **20. Community Acknowledgments**

We would like to thank the many communities of Alaska and their members who have helped contribute to this body of work. The knowledge, experiences, and insights have been instrumental in expanding the scope of our science and knowledge to encompass the many issues that face this important ecosystem. We appreciate feedback from those residing in the region that are willing to share their insights and participation in an open dialog about how we can improve our collective knowledge of the ecosystem and the region.

# **21. Technical Acknowledgments**

This quarto book is based off the NOAA-quarto-book GitHub repo designed by Eli Holmes.

This repo and GitHub Action was based on the tutorial by Openscapes quarto-website-tutorial by Julia Lowndes and Stefanie Butland.

## **21.1. Partners**

Scientists from the Alaska Fisheries Science Center conduct these bottom trawl surveys with participation from the Alaska Department of Fish & Game (ADF&G), the International Pacific Halibut Commission (IPHC), and universities. This research is conducted on chartered fishing vessels.

## 22. References

- Alaska Fisheries Information Network (AKFIN). (2023). *AFSC groundfish assessment program design-based production data*. NOAA Fisheries Alaska Fisheries Science Center, Goundfish Assessment Program; <https://www.psmfc.org/program/alaska-fisheries-information-network-akfin>; U.S. Dep. Commer.
- Hoff, G. R. (2016). *Results of the 2016 eastern Bering Sea upper continental slope survey of groundfishes and invertebrate resources* (NOAA Tech. Memo. NOAA-AFSC-339). U.S. Dep. Commer. <https://doi.org/10.7289/V5/TM-AFSC-339>
- Markowitz, E. H., Dawson, E. J., Anderson, A. B., Rohan, S. K., Charriere, N. E., Prohaska, B. K., and Stevenson, D. E. (2023). *Results of the 2022 eastern and northern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate fauna* (NOAA Tech. Memo. NMFS-AFSC-469; p. 213). U.S. Dep. Commer.
- Markowitz, E. H., Dawson, E. J., Anderson, C., Charriere, N. E., Richar, J. I., Rohan, S. K., Prohaska, B. K., Haehn, R. A., and Stevenson, D. E. (2022). *2022 northern Bering Sea groundfish and crab trawl survey highlights [Outreach]*. University of Alaska Fairbanks Strait Science Seminar; <https://www.youtube.com/watch?v=TGXN2pIDhfc>.
- NOAA Fisheries Alaska Fisheries Science Center. (2023). *Fisheries one stop shop public data: RACE division bottom trawl survey data query*. <https://www.fisheries.noaa.gov/foss>; U.S. Dep. Commer.
- NOAA Fisheries Alaska Fisheries Science Center, Goundfish Assessment Program. (2023). *AFSC groundfish assessment program design-based production data*. <https://www.fisheries.noaa.gov/alaska/science-data/groundfish-assessment-program-bottom-trawl-surveys>; U.S. Dep. Commer.
- Von Szalay, P. G., and Raring, N. W. (2018). *Data report: 2017 Gulf of Alaska bottom trawl survey* (NOAA Tech. Memo. NMFS-AFSC-374). U.S. Dep. Commer. <https://doi.org/10.7289/V5/TM-AFSC-374>
- Von Szalay, P. G., and Raring, N. W. (2020). *Data report: 2018 Aleutian Islands bottom trawl survey* (NOAA Tech. Memo. NMFS-AFSC-409). U.S. Dep. Commer. <https://doi.org/10.25923/qe5v-fz70>