GAP Production Data Documentation

Bering Sea Survey Team

Gulf of Alaska and Aleutian Island Survey Team

Table of contents

# 1. Survey Background

## 1.1 What we do

## 1.2 Who is conducting the research?

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

## 1.3 What is the research objective?

Learn more about the [program](https://www.fisheries.noaa.gov/alaska/science-data/groundfish-assessment-program-bottom-trawl-surveys). The objectives of these surveys are to:

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

## 1.4 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.

## 1.5 Bottom trawl surveys and regions



Each survey conducted by the [Groundfish Assessment Program](https://www.fisheries.noaa.gov/alaska/population-assessments/north-pacific-groundfish-stock-assessments-and-fishery-evaluation) are multispecies bottom trawl surveys. We collect environmental and biological data to assess how climate variability and [loss of sea](https://www.fisheries.noaa.gov/alaska/ecosystems/habitat-and-ecological-processes-research-regarding-loss-sea-ice) 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](https://www.fisheries.noaa.gov/alaska/science-data/alaska-age-and-growth-procedures-otolith-examination) to [determine fish ages](https://www.fisheries.noaa.gov/alaska/science-data/fish-otolith-chronologies). We use this information in [annual stock assessments](https://www.fisheries.noaa.gov/alaska/population-assessments/north-pacific-groundfish-stock-assessments-and-fishery-evaluation) and to assess the state of the ecosystem. This research is conducted on fishing industry contract vessels.

| **Survey** | **Survey Definition ID** | **Years** | **Depth (m)** | **Area (km2)** | **# 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 - 1,200 | 32,861.3 | 37 |  |
| Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey | 98 | 2023 - 1982 (41) | 1 - 200 | 492,989.9 | 28 | 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 | 2023 - 2010 (6) | 1 - 100 | 198,866.8 | 4 | 144 |

### 1.5.1 **Aleutian Islands**

(Von Szalay et al., 2023)

* 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](https://www.fisheries.noaa.gov/species/pacific-ocean-perch), [walleye pollock](https://www.fisheries.noaa.gov/species/alaska-pollock), [Pacific cod](https://www.fisheries.noaa.gov/species/pacific-cod), [sablefish](https://www.fisheries.noaa.gov/species/sablefish), and other rockfish species.

### 1.5.2 **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](https://www.fisheries.noaa.gov/species/pacific-ocean-perch), [walleye pollock](https://www.fisheries.noaa.gov/species/alaska-pollock), [Pacific cod](https://www.fisheries.noaa.gov/species/pacific-cod), flatfish, and other rockfish species.

### 1.5.3 **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](https://www.fisheries.noaa.gov/species/alaska-pollock), [Pacific cod](https://www.fisheries.noaa.gov/species/pacific-cod), [Greenland turbot](https://www.fisheries.noaa.gov/species/greenland-turbot), [yellowfin sole](https://www.fisheries.noaa.gov/species/yellowfin-sole), [northern rock sole](https://www.fisheries.noaa.gov/species/rock-sole), [red king crab](https://www.fisheries.noaa.gov/species/red-king-crab), and [snow](https://www.fisheries.noaa.gov/species/alaska-snow-crab) and Tanner crabs.

|  |
| --- |
| Strata used in the Eastern Bering Sea Survey. |

### 1.5.4 **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.

### 1.5.5 **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](https://www.fisheries.noaa.gov/species/pacific-ocean-perch), popeye grenadier, [walleye pollock](https://www.fisheries.noaa.gov/species/alaska-pollock), and [arrowtooth flounder](https://www.fisheries.noaa.gov/species/arrowtooth-flounder).

|  |
| --- |
| Strata used in the Bering Sea Slope Survey. |

# 2. Workflow

## 2.1 Operational Product Development Timeline

Over the course of the year, the survey team is developing a variety of different data products. Planning and preparation for surveys happens in the late winter and spring, surveys occur in the summer, data validation takes place over the course of the survey and after the survey, and data products are produced through fall and late winter.

|  | January | February | March | April | May | June | July | August | September | October | November | December |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Surveys |  |  |  |  | 1 | 1 | 1 | 1 |  |  |  |  |
| Planning | 1 | 1 | 1 |  |  |  |  |  | 1 | 1 | 1 | 1 |
| Development | 1 | 1 | 1 | 1 | 1 |  |  |  |  | 1 | 1 | 1 |
| Deployment (survey deliverables) |  |  |  |  |  |  |  |  | 1 | 1 | 1 | 1 |
| Deployment (survey operations) |  |  |  |  | 1 | 1 | 1 | 1 |  |  |  |  |
| Triage (fixing bugs and errors) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| User feedback and brainstorming | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

## 2.2 Data workflow from boat to production

Organisms first need to be collected aboard the vessel before data can be entered into tablets.

|  |
| --- |
| Figure 2.1: Simplified boat deck processing workflow. |

The objective of this process is to take raw data, QA/QC and clean these data, curate standard data products for these survey. Please note, through this process we are not providing “data” (what we consider lower level data material; see the data levels section below) but “data products”, which is intended to facilitate the most fool-proof standard interpretation of the data. These data products only use data from standard and validated hauls, and has undergone careful review.

**Once survey data collected on the vessel has been checked and validated**, the [gap\_products/code/run.R](https://github.com/afsc-gap-products/gap_products/blob/main/code/run.R) script is used to orchestrate a 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.

The data from these surveys constitute a **living data set** so we can continue to **provide the best available data to all partners, stakeholders, and fellow scientists**.

|  |
| --- |
| Figure 2.2: Simplified data workflow from boat to production. |

During each data product run cycle:

1. Versions of the tables in GAP\_PRODUCTS are locally imported 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.

|  |
| --- |
| Figure 2.3: Major end-users of the GAP data product tables. |

## 2.3 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 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](https://ladsweb.modaps.eosdis.nasa.gov/search/) 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](https://docs.google.com/presentation/d/1rWSZpeghWJqzWMIa5oBc4BCoy-zy1Yue86RoTw58u6M/edit?usp=sharing) on the topic.

* **Level 0**: Raw and unprocessed data. Ex: Data on the G drive, some tables in RACE\_DATA
* **Level 1**: 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 News/change logs

– [GAP\_PRODUCTS ChangeLog (last produced on 2024-01-09) using gapindex v2.1.3](https://raw.githubusercontent.com/afsc-gap-products/gap_products/main/content/intro-news/2024-01-09.txt): gapindex Version

– [GAP\_PRODUCTS ChangeLog (last produced on 2023-11-17) using gapindex v2.1.2](https://raw.githubusercontent.com/afsc-gap-products/gap_products/main/content/intro-news/2023-11-17.txt): A new version of gapindex ([v2.1.2](https://github.com/afsc-gap-products/gapindex/releases/tag/v2.1.2)) was used to produced these data. There was a slight change to how subarea biomass totals are calculated that was not fully addressed in v2.1.1. The modified biomass records reflect this change.

– [GAP\_PRODUCTS ChangeLog (last produced on 2023-11-14) using gapindex v2.1.1](https://raw.githubusercontent.com/afsc-gap-products/gap_products/main/content/intro-news/2023-11-14_v2-1-1.txt): A new version of gapindex (v2.1.1) was used to produced these data. There was a slight change to how subarea biomass totals are calculated. The modified biomass records reflect this change. New 2022 otolith data were available since the last iteration of the GAP\_PRODUCTS for Aleutian Island Pacific ocean perch and northern rockifsh and Eastern Bering Sea northern rock sole. Zero-filled CPUE records for four GOA species codes (SPECIES\_CODE: 21210, 30010, 30360, 77102, 98101) were added due to how the 1990 data were integrated in the last production run of GAP\_PRODUCTS. Two Arctic cod (SPECIES\_CODE: 21725) and one plain sculpin (SPECIES\_CODE: 21371) count records were modified in the NBS data, which changes the numerical CPUE estimates for those hauls which changes the estimated population abundance and size composition for those species.

– [Groundfish Assessment Program Survey Data Serving and Data Improvements](https://raw.githubusercontent.com/afsc-gap-products/gap_products/main/content/intro-news/2023-06-20_datachangesbreif): Initial data changes brief distributed to SSMA and other partners by Ned Laman, Zack Oyafuso, and Emily Markowitz

– [Run 2023-06-01 gapindex v2.1.0](https://raw.githubusercontent.com/afsc-gap-products/gap_products/main/content/intro-news/2023-06-01_v2-1-0.txt): Initial compiling and planning notes

# 4. Code of Conduct

## 4.1 What are Codes of Conduct?

Codes of Conduct are voluntary sets of rules that assist creators, developers, and users of code and data with data protection compliance and accountability in specific sectors or relating to particular processing operations.

Codes can help organizations to ensure all participants follow best practices and rules designed specifically for their sector or processing operations, thus enhancing compliance and collaboration. They are developed and managed by an association or other body (the ‘Code Owner’) which is representative of a sector (or category of data controllers or processors), with the expert and sectoral knowledge of how to enhance data protection in their area.

### 4.1.1 [Code of Conduct](https://github.com/nmfs-opensci/.github/blob/main/CODE_OF_CONDUCT.md) from the [nmfs-opensci GitHub](https://nmfs-opensci.github.io/).

# 5. NOAA Fisheries Open Science Code of Conduct

This code of conduct was developed and adapted from the Atom code of conduct in October 2021.

## 5.1 Our Pledge

In the interest of fostering an open and welcoming environment, we as contributors and maintainers pledge to making participation in our project and our community a harassment-free experience for everyone, regardless of age, body size, disability, ethnicity, gender identity and expression, level of experience, nationality, personal appearance, race, religion, or sexual identity and orientation.

## 5.2 Our Standards

Examples of behavior that contributes to creating a positive environment include:

* Using welcoming and inclusive language
* Being respectful of differing viewpoints and experiences
* Gracefully accepting constructive criticism
* Focusing on what is best for the community
* Showing empathy towards other community members

Examples of unacceptable behavior by participants include:

* The use of sexualized language or imagery and unwelcome sexual attention or advances
* Trolling, insulting/derogatory comments, and personal or political attacks
* Public or private harassment
* Publishing others’ private information, such as a physical or electronic address, without explicit permission
* Other conduct which could reasonably be considered inappropriate in a professional setting

## 5.3 Our Responsibilities

Project maintainers are responsible for clarifying the standards of acceptable behavior and are expected to take appropriate and fair corrective action in response to any instances of unacceptable behavior.

Project maintainers have the right and responsibility to remove, edit, or reject comments, commits, code, wiki edits, issues, and other contributions that are not aligned to this Code of Conduct, or to ban temporarily or permanently any contributor for other behaviors that they deem inappropriate, threatening, offensive, or harmful.

## 5.4 Scope

This Code of Conduct applies both within project spaces and in public spaces when an individual is representing the project or its community. Examples of representing a project or community include using an official project e-mail address, posting via an official social media account, or acting as an appointed representative at an online or offline event. Representation of a project may be further defined and clarified by project maintainers.

## 5.5 Enforcement

Instances of abusive, harassing, or otherwise unacceptable behavior may be reported by contacting the project team. All complaints will be reviewed and investigated and will result in a response that is deemed necessary and appropriate to the circumstances. Further details of specific enforcement policies may be posted separately.

## 5.6 Attribution

This Code of Conduct is adapted from the [Contributor Covenant](https://contributor-covenant.org), version 1.4, available at [https://contributor-covenant.org/version/1/4](https://contributor-covenant.org/version/1/4/)

# 6. Data description

## 6.1 Data tables

### 6.1.1 AGECOMP

Region-level age compositions by sex/length bin.

Number of rows: 547,436

Number of columns: 9

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AGE

Taxon age bin (yrs)

integer

NUMBER(38,0)

Age bin of taxon. 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 (millimeters)

LENGTH\_MM\_SD

Standard deviation of length at age weighted by numbers at length

numeric

NUMBER(38,3)

Variance of mean length.

POPULATION\_COUNT

Estimated population

numeric

NUMBER(38,0)

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

### 6.1.2 AREA

This is a table

Number of rows: 511

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 (km2)

kilometers squared

NUMBER(38,3)

Area in 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.

AREA\_TYPE

Area ID type description

category

VARCHAR2(255 BYTE)

The type of stratum that AREA\_ID represents. Types include: STRATUM (the smallest building-block unit of area in these surveys), REGION, DEPTH, SUBAREA, INPFC BY DEPTH, INPFC, SUBAREA BY DEPTH, REGULATORY AREA, NMFS STATISTICAL AREA.

CRS

Coordinate reference system

ID code

VARCHAR2(255 BYTE)

The coordinate reference system (CRS) that shapefiles were created in or areas (like AREA\_KM2) are calculated in, as defined by https://spatialreference.org/ (e.g., “+proj=longlat”, “EPSG:3338”).

DEPTH\_MAX\_M

Area ID maximum depth (m)

meters

NUMBER(38,3)

Maximum depth (meters).

DEPTH\_MIN\_M

Area ID minimum depth (m)

meters

NUMBER(38,3)

Minimum depth (meters).

DESCRIPTION

Description

text

VARCHAR2(4000 BYTE)

Description of row observation.

DESIGN\_YEAR

Design year

year

NUMBER(10,0)

Year ID associated with a given value AREA\_ID. This field describes the changes in the survey design over time.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

### 6.1.3 BIOMASS

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

Number of rows: 4,714,192

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

Estimated biomass

numeric

NUMBER(38,6)

The estimated total biomass.

BIOMASS\_VAR

Estimated biomass variance

numeric

NUMBER(38,6)

The estimated variance associated with the total biomass.

CPUE\_KGKM2\_MEAN

Mean weight CPUE

kilograms per kilometers squared

NUMBER(38,6)

The mean catch weight (kilograms) per unit effort (area swept by the net, units squared kilometers).

CPUE\_KGKM2\_VAR

Variance of the mean weight CPUE

kilograms per kilometers squared

NUMBER(38,6)

The variance of mean catch weight (kilograms) per unit effort (area swept by the net, units squared kilometers).

CPUE\_NOKM2\_MEAN

Mean numeric CPUE

count per kilometers squared

NUMBER(38,6)

The mean of numerical catch per unit effort (area swept by the net, units square kilometers).

CPUE\_NOKM2\_VAR

Variance of the mean numeric CPUE

count per kilometers squared

NUMBER(38,6)

The variance of mean numerical catch per unit effort (area swept by the net, units square kilometers).

N\_COUNT

Hauls with taxon counts

numeric

NUMBER(38,0)

Total number of hauls with positive count data.

N\_HAUL

Valid hauls

count

NUMBER(38,0)

Total number of hauls.

N\_LENGTH

Hauls with taxon lengths

count

NUMBER(38,0)

Total number of hauls with length data.

N\_WEIGHT

Hauls with catch

count

NUMBER(38,0)

Total number of hauls with positive catch biomass.

POPULATION\_COUNT

Estimated population

numeric

NUMBER(38,0)

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

### 6.1.4 CPUE

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

Number of rows: 39,016,899

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 whole number of individuals caught in haul or samples collected.

CPUE\_KGKM2

Weight CPUE (kg/km2)

kilograms per kilometers squared

NUMBER(38,6)

Catch weight (kilograms) per unit effort (area swept by the net, units square kilometers).

CPUE\_NOKM2

Number CPUE (no/km2)

count per kilometers squared

NUMBER(38,6)

Numerical catch per unit effort (area swept by the net, units square kilometers).

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

WEIGHT\_KG

Sample or taxon weight (kg)

kilograms

NUMBER(38,3)

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

### 6.1.5 SURVEY\_DESIGN

This is a 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)

Year ID associated with a given value AREA\_ID. This field describes the changes in the survey design over time.

SURVEY

Survey Name

text

VARCHAR2(255 BYTE)

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

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

### 6.1.6 METADATA\_TABLE

These columns provide the table metadata for all of the tables and views in GAP\_PRODUCTS. These tables are created by the Resource Assessment and Conservation Engineering Division (RACE) Groundfish Assessment Program (GAP) of the Alaska Fisheries Science Center (AFSC). The GitHub repository for the scripts that created this code can be found at https://github.com/afsc-gap-products/gap\_products. These data were last updated February 10, 2024. There are no legal restrictions on access to the data. For more information about codes used in the tables, please refer to the survey code books (https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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(4000 BYTE)

Table metadata sentence.

METADATA\_SENTENCE\_NAME

Metadata sentence name

text

VARCHAR2(4000 BYTE)

Name of table metadata sentence.

METADATA\_SENTENCE\_TYPE

Sentence type

text

VARCHAR2(4000 BYTE)

Type of sentence to have in table metadata.

### 6.1.7 STRATUM\_GROUPS

This is a 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)

Year ID associated with a given value AREA\_ID. This field describes the changes in the survey design over time.

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 region. Stratum of value 0 indicates experimental tows.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

### 6.1.8 SIZECOMP

Stratum/subarea/region-level size compositions by sex. This table was 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. For more information about codes used in the tables, please refer to the survey code books (https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual). These data were last updated January 10, 2024.

Number of rows: 3,180,973

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)

Length bin in millimeters.

POPULATION\_COUNT

Estimated population

numeric

NUMBER(38,0)

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

# 7. Data description

[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.

## 7.1 Data tables

### 7.1.1 AKFIN\_AGECOMP

This table is a copy of GAP\_PRODUCTS.AGECOMP and does not have any other object dependencies. These data are produced 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. These data were last updated February 10, 2024.

Number of rows: 547,436

Number of columns: 9

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AGE

Taxon age bin (yrs)

integer

NUMBER(38,0)

Age bin of taxon. 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 (millimeters)

LENGTH\_MM\_SD

Standard deviation of length at age weighted by numbers at length

numeric

NUMBER(38,3)

Variance of mean length.

POPULATION\_COUNT

Estimated population

numeric

NUMBER(38,0)

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

### 7.1.2 AKFIN\_AREA

This table is a copy of GAP\_PRODUCTS.AREA and does not have any other object dependencies. These data are produced 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. These data were last updated February 10, 2024.

Number of rows: 511

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 (km2)

kilometers squared

NUMBER(38,3)

Area in 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.

AREA\_TYPE

Area ID type description

category

VARCHAR2(255 BYTE)

The type of stratum that AREA\_ID represents. Types include: STRATUM (the smallest building-block unit of area in these surveys), REGION, DEPTH, SUBAREA, INPFC BY DEPTH, INPFC, SUBAREA BY DEPTH, REGULATORY AREA, NMFS STATISTICAL AREA.

CRS

Coordinate reference system

ID code

VARCHAR2(255 BYTE)

The coordinate reference system (CRS) that shapefiles were created in or areas (like AREA\_KM2) are calculated in, as defined by https://spatialreference.org/ (e.g., “+proj=longlat”, “EPSG:3338”).

DEPTH\_MAX\_M

Area ID maximum depth (m)

meters

NUMBER(38,3)

Maximum depth (meters).

DEPTH\_MIN\_M

Area ID minimum depth (m)

meters

NUMBER(38,3)

Minimum depth (meters).

DESCRIPTION

Description

text

VARCHAR2(4000 BYTE)

Description of row observation.

DESIGN\_YEAR

Design year

year

NUMBER(10,0)

Year ID associated with a given value AREA\_ID. This field describes the changes in the survey design over time.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

### 7.1.3 AKFIN\_BIOMASS

This table is a copy of GAP\_PRODUCTS.BIOMASS and does not have any other object dependencies. These data are produced 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. These data were last updated February 10, 2024.

Number of rows: 4,714,192

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

Estimated biomass

numeric

NUMBER(38,6)

The estimated total biomass.

BIOMASS\_VAR

Estimated biomass variance

numeric

NUMBER(38,6)

The estimated variance associated with the total biomass.

CPUE\_KGKM2\_MEAN

Mean weight CPUE

kilograms per kilometers squared

NUMBER(38,6)

The mean catch weight (kilograms) per unit effort (area swept by the net, units squared kilometers).

CPUE\_KGKM2\_VAR

Variance of the mean weight CPUE

kilograms per kilometers squared

NUMBER(38,6)

The variance of mean catch weight (kilograms) per unit effort (area swept by the net, units squared kilometers).

CPUE\_NOKM2\_MEAN

Mean numeric CPUE

count per kilometers squared

NUMBER(38,6)

The mean of numerical catch per unit effort (area swept by the net, units square kilometers).

CPUE\_NOKM2\_VAR

Variance of the mean numeric CPUE

count per kilometers squared

NUMBER(38,6)

The variance of mean numerical catch per unit effort (area swept by the net, units square kilometers).

N\_COUNT

Hauls with taxon counts

numeric

NUMBER(38,0)

Total number of hauls with positive count data.

N\_HAUL

Valid hauls

count

NUMBER(38,0)

Total number of hauls.

N\_LENGTH

Hauls with taxon lengths

count

NUMBER(38,0)

Total number of hauls with length data.

N\_WEIGHT

Hauls with catch

count

NUMBER(38,0)

Total number of hauls with positive catch biomass.

POPULATION\_COUNT

Estimated population

numeric

NUMBER(38,0)

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

### 7.1.4 AKFIN\_CATCH

snapshot table for snapshot GAP\_PRODUCTS.AKFIN\_CATCH

Number of rows: 989,351

Number of columns: 6

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

CATCHJOIN

Catch observation ID

ID code

NUMBER(38,0)

Unique integer ID assigned to each survey, vessel, year, and catch observation combination.

COUNT

Taxon count

count, whole number resolution

NUMBER(38,0)

Total whole number of individuals caught in haul or samples collected.

CRUISEJOIN

Cruise ID

ID code

NUMBER(38,0)

Unique integer ID assigned to each survey, vessel, and year combination.

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

WEIGHT\_KG

Sample or taxon weight (kg)

kilograms

NUMBER(38,3)

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

### 7.1.5 AKFIN\_CPUE

This table is a copy of GAP\_PRODUCTS.CPUE and does not have any other object dependencies. These data are produced 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. These data were last updated February 10, 2024.

Number of rows: 39,016,899

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 whole number of individuals caught in haul or samples collected.

CPUE\_KGKM2

Weight CPUE (kg/km2)

kilograms per kilometers squared

NUMBER(38,6)

Catch weight (kilograms) per unit effort (area swept by the net, units square kilometers).

CPUE\_NOKM2

Number CPUE (no/km2)

count per kilometers squared

NUMBER(38,6)

Numerical catch per unit effort (area swept by the net, units square kilometers).

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

WEIGHT\_KG

Sample or taxon weight (kg)

kilograms

NUMBER(38,3)

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

### 7.1.6 AKFIN\_CRUISE

This is the cruise data table. These data are produced 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. These data were last updated February 10, 2024.

Number of rows: 187

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 integer 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)

Unique integer ID 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)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SURVEY\_NAME

Survey name

text

VARCHAR2(255 BYTE)

Long name of the survey conducted

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

### 7.1.7 AKFIN\_HAUL

snapshot table for snapshot GAP\_PRODUCTS.AKFIN\_HAUL

Number of rows: 36,114

Number of columns: 25

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

ACCESSORIES

Type of gear accessories used on the net

ID code

NUMBER(38,0)

Type of accessories used on net. For a complete list of accessories ID codes, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

BOTTOM\_TYPE

Seafloor bottom type code

ID code

NUMBER(38,0)

Bottom type on sea floor at haul location. For a complete list of bottom type ID codes, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

CRUISEJOIN

Cruise ID

ID code

NUMBER(38,0)

Unique integer ID assigned to each survey, vessel, and year combination.

DATE\_TIME\_START

Start date and time

MM/DD/YYYY HH::MM

TIMESTAMP

The date (MM/DD/YYYY) and time (HH:MM) of the beginning of the haul. All dates and times are in Alaska time (AKDT) of Anchorage, AK, USA (UTC/GMT -8 hours).

DEPTH\_GEAR\_M

Depth of gear (m)

degrees Celsius

NUMBER(38,1)

Depth of gear (meters).

DEPTH\_M

Depth (m)

degrees Celsius

NUMBER(38,1)

Bottom depth (meters).

DISTANCE\_FISHED\_KM

Distance fished (km)

degrees Celsius

NUMBER(38,3)

Distance the net fished (thousands 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

Type of gear used on the net

ID code

NUMBER(38,0)

Type of gear used on net. For a complete list of gear ID codes, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

GEAR\_TEMPERATURE\_C

Gear temperature (Degrees Celsius)

degrees Celsius

NUMBER(38,1)

Temperature recorded by net gear (tenths of a degree Celsius); NA indicates removed or missing values.

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

LATITUDE\_DD\_END

End latitude (decimal degrees)

decimal degrees

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\_MEASURED

Net measured during haul

logical

BINARY\_DOUBLE

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

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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 region. 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.

WIRE\_LENGTH\_M

Trawl wire length

meters

NUMBER(38,0)

Length of wire deployed during a given haul in meters.

### 7.1.8 AKFIN\_LENGTH

snapshot table for snapshot GAP\_PRODUCTS.AKFIN\_LENGTH

Number of rows: 4,601,070

Number of columns: 7

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

FREQUENCY

Count of observation

count

NUMBER(38,0)

Frequency, or count, of an observation.

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 bin in millimeters.

LENGTH\_TYPE

Length type

ID code

NUMBER(38,0)

How the taxon was measured (e.g., fork length, carapace width). For a complete list of length\_type ID codes, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SAMPLE\_TYPE

Sample type

ID code

NUMBER(38,0)

Sampling information on how the taxon was sampled. For a complete list of length\_type ID codes, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

### 7.1.9 AKFIN\_METADATA\_COLUMN

This table is a copy of GAP\_PRODUCTS.METADATA\_COLUMN and does not have any other object dependencies. These data are produced 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. These data were last updated February 10, 2024.

Number of rows: 158

Number of columns: 5

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

METADATA\_COLNAME

Column name

text

VARCHAR2(4000 BYTE)

Name of the column in a table.

METADATA\_COLNAME\_DESC

column description

text

VARCHAR2(4000 BYTE)

Description of the column.

METADATA\_COLNAME\_LONG

Column name spelled out

text

VARCHAR2(4000 BYTE)

Long name for the column.

METADATA\_DATATYPE

Oracle datatype code

text

VARCHAR2(4000 BYTE)

Oracle data type of data column.

METADATA\_UNITS

Units

category

VARCHAR2(4000 BYTE)

Units of the column.

### 7.1.10 AKFIN\_SIZECOMP

This table is a copy of GAP\_PRODUCTS.SIZECOMP and does not have any other object dependencies. These data are produced 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. These data were last updated February 10, 2024.

Number of rows: 3,180,973

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)

Length bin in millimeters.

POPULATION\_COUNT

Estimated population

numeric

NUMBER(38,0)

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

### 7.1.11 AKFIN\_SPECIMEN

snapshot table for snapshot GAP\_PRODUCTS.AKFIN\_SPECIMEN

Number of rows: 634,835

Number of columns: 12

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

AGE

Taxon age bin (yrs)

integer

NUMBER(38,0)

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

AGE\_DETERMINATION\_METHOD

Aging method

ID code

NUMBER(10,0)

Numeric code corresponding to the method of age determination. For a complete list of age determination codes, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

GONAD\_G

Weight of gonads (g)

grams

NUMBER(38,1)

Weight of specimen gonads (grams).

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 bin in millimeters.

MATURITY

Specimen maturity code

ID code

NUMBER(38,0)

The maturity code or the condition identified by the maturity code.

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SPECIMEN\_ID

Specimen unique ID

ID code

NUMBER(38,0)

Each individual examined must have a number assigned to it that is unique within each haul (0001 to 9999), though specimen numbers may be repeated between hauls

SPECIMEN\_SAMPLE\_TYPE

Specimen sample type

ID code

NUMBER(38,0)

The specimen sample type ID code as defined in the RACE\_DATA.SPECIMEN\_SAMPLE\_TYPES table. For a complete list of Specimen sample type ID codes, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SPECIMEN\_SUBSAMPLE\_METHOD

Specimen subsample method

ID code

NUMBER(38,0)

For a complete list of specimen subsample method ID codes, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

WEIGHT\_G

Specimen weight (g)

grams

NUMBER(38,1)

Weight of specimen (grams).

### 7.1.12 AKFIN\_STRATUM\_GROUPS

This table is a copy of GAP\_PRODUCTS.STRATUM\_GROUPS and does not have any other object dependencies. These data are produced 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. These data were last updated February 10, 2024.

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)

Year ID associated with a given value AREA\_ID. This field describes the changes in the survey design over time.

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 region. Stratum of value 0 indicates experimental tows.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

### 7.1.13 AKFIN\_SURVEY\_DESIGN

This table is a copy of GAP\_PRODUCTS.SURVEY\_DESIGN and does not have any other object dependencies. These data are produced 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. These data were last updated February 10, 2024.

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)

Year ID associated with a given value AREA\_ID. This field describes the changes in the survey design over time.

SURVEY

Survey Name

text

VARCHAR2(255 BYTE)

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

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

### 7.1.14 AKFIN\_TAXONOMIC\_CLASSIFICATION

NAThese data are produced 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. These data were last updated February 10, 2024.

Number of rows: 2,699

Number of columns: 19

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

CLASS\_TAXON

Class phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of class 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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

DATABASE

Database source

category

VARCHAR2(255 BYTE)

Taxonomic database source, either ITIS or WoRMS.

DATABASE\_ID

Species ID in database

ID code

VARCHAR2(255 BYTE)

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

FAMILY\_TAXON

Family phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of family of a given species.

GENUS\_TAXON

Genus phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of genus of a given species.

ID\_RANK

Lowest taxonomic rank

text

VARCHAR2(255 BYTE)

Lowest taxonomic rank of a given species entry.

KINGDOM\_TAXON

Kingdom phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of kingdom of a given species.

ORDER\_TAXON

Order phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of order of a given species.

PHYLUM\_TAXON

Phylum phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of phylum 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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SPECIES\_NAME

Scientific name of species

text

VARCHAR2(255 BYTE)

Scientific name of species.

SUBCLASS\_TAXON

Subclass phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of subclass of a given species.

SUBFAMILY\_TAXON

Subfamily phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of subfamily of a given species.

SUBORDER\_TAXON

Suborder phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of suborder of a given species.

SUBPHYLUM\_TAXON

Subphylum phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of subphylum of a given species.

SUPERCLASS\_TAXON

Superclass phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of superclass of a given species.

SUPERFAMILY\_TAXON

Superfamily phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of superfamily of a given species.

SUPERORDER\_TAXON

Superorder phylogenetic rank

category

VARCHAR2(255 BYTE)

Phylogenetic latin rank of superorder of a given species.

# 8. Access data

## 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.

### 8.0.1 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 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.

After you connect to VPN, you’ll be able to log into Oracle.

library(RODBC)  
channel <- gapindex::get\_connected()

## Data SQL Query Examples:

library(gapindex)  
library(RODBC)  
library(flextable)  
library(ggplot2)  
library(magrittr)  
library(dplyr)

### 8.0.2 Ex. Select all data from tables

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.

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\_TAXONOMIC\_CLASSIFICATION"  
)  
  
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")))  
}

library(odbc)  
library(RODBC)  
library(dbplyr)  
  
my\_spp\_codes <- c(  
30010, # Sebastolobus sp. thornyhead unid.  
30020, # Sebastolobus alascanus shortspine thornyhead  
30025, # Sebastolobus macrochir broadfin thornyhead  
30330, # Sebastes melanops black rockfish  
30430, # Sebastes proriger redstripe rockfish  
30470, # Sebastes ruberrimus yelloweye rockfish  
30475, # Sebastes babcocki redbanded rockfish  
30535, # Sebastes variegatus harlequin rockfish  
30560, # Sebastes zacentrus sharpchin rockfish  
30600, # Sebastes reedi yellowmouth rockfish  
30030, # Sebastolobus altivelis longspine thornyhead  
30040, # Sebastes sp. rockfish unid.  
30100, # Sebastes brevispinis silvergray rockfish  
30150, # NA dusky and dark rockfishes unid.  
30152, # Sebastes variabilis dusky rockfish  
30170, # Sebastes crameri darkblotched rockfish  
30270) # Sebastes helvomaculatus rosethorn rockfish  
  
a <- dplyr::tbl(channel, dplyr::sql('gap\_products.akfin\_biomass')) %>%   
 dplyr::rename\_all(tolower) %>%   
 dplyr::select(survey\_definition\_id, area\_id, species\_code, year, biomass\_mt, biomass\_var) %>%   
 dplyr::filter(species\_code %in% my\_spp\_codes &   
 area\_id %in% 99904 &   
 year >= 1991) %>%   
 dplyr::collect()   
  
flextable::flextable(head(a)) %>%   
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra()

### 8.0.3 Ex. CPUE for all EBS and NBS stations with associated haul, cruise, and species information.

a <- RODBC::sqlQuery(channel = channel, # NOT RACEBASE.HAUL  
 query = paste0(  
 "  
-- Select columns for output data  
SELECT  
cr.CRUISEJOIN,  
cr.CRUISE,  
cr.YEAR,  
cr.SURVEY\_DEFINITION\_ID,  
cr.SURVEY\_NAME,  
cr.VESSEL\_ID,  
cr.VESSEL\_NAME,  
cp.HAULJOIN,  
cp.SPECIES\_CODE,  
tt.SPECIES\_NAME,  
tt.COMMON\_NAME,  
cp.WEIGHT\_KG,  
cp.COUNT,  
cp.AREA\_SWEPT\_KM2,  
cp.CPUE\_KGKM2,  
cp.CPUE\_NOKM2,  
hh.HAUL,  
hh.STATION  
  
-- Identify what tables to pull data from  
FROM GAP\_PRODUCTS.AKFIN\_HAUL hh  
LEFT JOIN GAP\_PRODUCTS.AKFIN\_CRUISE cr  
ON hh.CRUISEJOIN = cr.CRUISEJOIN  
LEFT JOIN GAP\_PRODUCTS.AKFIN\_CPUE cp  
ON hh.HAULJOIN = cp.HAULJOIN  
LEFT JOIN GAP\_PRODUCTS.TAXONOMIC\_CLASSIFICATION tt  
ON cp.SPECIES\_CODE = tt.SPECIES\_CODE  
  
-- Filter for EBS and NBS observations  
WHERE SURVEY\_DEFINITION\_ID IN (143, 98) -- 143 NBS, 98 EBS  
AND tt.SURVEY\_SPECIES = 1  
  
-- Only return the first 3 rows because otherwise this would be a huge table!  
FETCH FIRST 3 ROWS ONLY;"))   
  
flextable::flextable(head(a[,2:8])) %>%   
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra()

| **CRUISE** | **YEAR** | **SURVEY\_DEFINITION\_ID** | **SURVEY\_NAME** | **VESSEL\_ID** | **VESSEL\_NAME** | **HAULJOIN** |
| --- | --- | --- | --- | --- | --- | --- |
| 198,203 | 1,982 | 98 | Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey | 1 | CHAPMAN | 877 |
| 198,203 | 1,982 | 98 | Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey | 1 | CHAPMAN | 877 |
| 198,203 | 1,982 | 98 | Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey | 1 | CHAPMAN | 877 |

### 8.0.4 Ex. CPUE for all stations contained in the INPFC Shumagin region (AREA\_ID = 919) for Pacific cod.

dat <- RODBC::sqlQuery(channel = channel,  
 query =  
 "  
-- Select columns for output data  
SELECT   
HAULJOIN,   
SPECIES\_CODE,   
STRATUM,   
LATITUDE\_DD\_START,   
LONGITUDE\_DD\_START,  
CPUE\_KGKM2,   
GEAR\_TEMPERATURE\_C  
  
-- Identify what tables to pull data from  
FROM GAP\_PRODUCTS.AKFIN\_CPUE cpue  
LEFT JOIN GAP\_PRODUCTS.AKFIN\_HAUL haul  
USING (HAULJOIN)   
  
-- Filter for P. Cod observations  
WHERE SPECIES\_CODE IN (21720)  
  
-- Select all stratum within the area\_id 919 (INPFC Shumagin region)  
AND haul.STRATUM IN  
(  
SELECT   
STRATUM  
FROM GAP\_PRODUCTS.AKFIN\_STRATUM\_GROUPS   
WHERE AREA\_ID = 919  
);")

dat <- dat %>%   
 dplyr::select(HAULJOIN, STRATUM, SPECIES\_CODE, LATITUDE\_DD\_START, LONGITUDE\_DD\_START, CPUE\_KGKM2, GEAR\_TEMPERATURE\_C) %>%   
 dplyr::mutate(SPECIES\_CODE = as.character(SPECIES\_CODE),   
 STRATUM = as.character(STRATUM)) %>%   
 dplyr::arrange(SPECIES\_CODE)  
  
flextable::flextable(head(dat)) %>%   
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra()

| **HAULJOIN** | **STRATUM** | **SPECIES\_CODE** | **LATITUDE\_DD\_START** | **LONGITUDE\_DD\_START** | **CPUE\_KGKM2** | **GEAR\_TEMPERATURE\_C** |
| --- | --- | --- | --- | --- | --- | --- |
| -22,147 | 11 | 21720 | 53.92966 | -164.6276 | 105.7733 | 4.3 |
| -22,119 | 111 | 21720 | 53.79215 | -165.3128 | 451.5863 | 4.5 |
| -22,107 | 12 | 21720 | 55.49881 | -161.5752 | 1,163.0550 | 5.2 |
| -22,099 | 12 | 21720 | 54.85306 | -162.9025 | 1,301.4142 | 4.5 |
| -22,153 | 11 | 21720 | 54.08538 | -163.3250 | 724.5604 | 4.6 |
| -22,135 | 11 | 21720 | 54.42703 | -162.4245 | 210.1829 | 4.9 |

### 8.0.5 Ex. EBS Pacific Ocean perch CPUE and [akgfmaps](https://github.com/afsc-gap-products/akgfmaps) map

Pacific Ocean perch catch-per-unit-effort estimates for EBS in 2021 from GAP\_PRODUCTS.AKFIN\_CPUE and map constructed using [akgfmaps](https://github.com/afsc-gap-products/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 columns for output data  
SELECT   
(cp.CPUE\_KGKM2/100) CPUE\_KGHA, -- akgfmaps is expecting hectares  
hh.LATITUDE\_DD\_START LATITUDE,  
hh.LONGITUDE\_DD\_START LONGITUDE  
  
-- Use HAUL data to obtain LATITUDE & LONGITUDE and connect to cruisejoin  
FROM GAP\_PRODUCTS.AKFIN\_CPUE cp  
LEFT JOIN GAP\_PRODUCTS.AKFIN\_HAUL hh  
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  
  
-- Filter data  
WHERE cp.SPECIES\_CODE = 30060   
AND cc.SURVEY\_DEFINITION\_ID = 98   
AND cc.YEAR = 2021;")

dat %>%   
 dplyr::arrange(desc(CPUE\_KGHA)) %>%   
 head() %>%   
 flextable::flextable() %>%   
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra()

| **CPUE\_KGHA** | **LATITUDE** | **LONGITUDE** |
| --- | --- | --- |
| 10.1768965 | 57.64871 | -173.3735 |
| 6.2734470 | 56.36952 | -169.4604 |
| 3.0252034 | 56.66253 | -171.9549 |
| 1.8214628 | 57.98912 | -173.4816 |
| 0.5535672 | 55.65865 | -168.1804 |
| 0.2813533 | 57.32545 | -173.3217 |

# 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]  
[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)

|  |
| --- |
| Ex. 6: EBS Pacific Ocean perch CPUE and [akgfmaps](https://github.com/afsc-gap-products/akgfmaps) map. |

### 8.0.6 Ex. 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 =   
 "  
-- Manipulate data to join to  
WITH FILTERED\_STRATA AS (  
SELECT AREA\_ID, DESCRIPTION FROM GAP\_PRODUCTS.AKFIN\_AREA  
WHERE AREA\_TYPE in ('REGULATORY\_AREA', 'REGION')   
AND SURVEY\_DEFINITION\_ID = 47)  
  
-- Select columns for output data  
SELECT   
BIOMASS\_MT,  
POPULATION\_COUNT,   
YEAR,   
DESCRIPTION  
  
-- Identify what tables to pull data from  
FROM GAP\_PRODUCTS.AKFIN\_BIOMASS BIOMASS  
JOIN FILTERED\_STRATA STRATA   
ON STRATA.AREA\_ID = BIOMASS.AREA\_ID  
  
-- Filter data results  
WHERE 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)) %>%   
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra() %>%  
 flextable::colformat\_num(x = ., j = "YEAR", big.mark = "")

| **BIOMASS\_MT** | **POPULATION\_COUNT** | **YEAR** | **DESCRIPTION** |
| --- | --- | --- | --- |
| 157,295.1 | 317,129,408 | 1990 | GOA Region: All Strata |
| 157,295.1 | 317,129,408 | 1990 | GOA Region: All Strata |
| 483,622.6 | 833,902,161 | 1993 | GOA Region: All Strata |
| 483,622.6 | 833,902,161 | 1993 | GOA Region: All Strata |
| 771,412.8 | 1,252,616,603 | 1996 | GOA Region: All Strata |
| 771,412.8 | 1,252,616,603 | 1996 | GOA Region: 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

|  |
| --- |
| Ex. 1: GOA Pacific Ocean perch biomass and abundance. |

### 8.0.7 Ex. 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](https://cran.r-project.org/web/packages/ggridges/vignettes/introduction.html).

dat <- RODBC::sqlQuery(channel = channel,   
 query =   
 "  
-- Manipulate data to join to  
WITH FILTERED\_STRATA AS (  
SELECT   
AREA\_ID,   
DESCRIPTION  
  
-- Identify what tables to pull data from  
FROM GAP\_PRODUCTS.AKFIN\_AREA  
WHERE AREA\_TYPE = 'REGION'   
AND SURVEY\_DEFINITION\_ID = 52)  
  
-- Select columns for output data  
SELECT   
LENGTH\_MM,   
YEAR  
FROM GAP\_PRODUCTS.AKFIN\_SIZECOMP SIZECOMP  
JOIN FILTERED\_STRATA STRATA   
ON STRATA.AREA\_ID = SIZECOMP.AREA\_ID  
  
-- Filter data results  
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) %>%   
 head() %>%   
 flextable::flextable() %>%   
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra() %>%  
 flextable::colformat\_num(x = ., j = "year", big.mark = "")  
dat0

| **length\_mm** | **year** | **length\_cm** |
| --- | --- | --- |
| 170 | 2018 | 17 |
| 180 | 2018 | 18 |
| 190 | 2018 | 19 |
| 200 | 2018 | 20 |
| 210 | 2018 | 21 |
| 220 | 2018 | 22 |

# install.packages("ggridges")  
library(ggridges)  
figure <-   
 ggplot2::ggplot(  
 data = dat,   
 mapping = aes(x = LENGTH\_MM, 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

|  |
| --- |
| Ex. 2: AI Rock sole size compositions and ridge plot. |

### 8.0.8 Ex. 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 =   
 "  
-- Manipulate data to join to  
WITH FILTERED\_STRATA AS (  
SELECT   
AREA\_ID,   
DESCRIPTION   
FROM GAP\_PRODUCTS.AKFIN\_AREA  
WHERE AREA\_TYPE = 'REGION' AND   
SURVEY\_DEFINITION\_ID = 98)  
  
-- Select columns for output data  
SELECT   
AGECOMP.AGE,   
AGECOMP.POPULATION\_COUNT,   
AGECOMP.SEX  
  
-- Identify what tables to pull data from  
FROM GAP\_PRODUCTS.AKFIN\_AGECOMP AGECOMP  
JOIN FILTERED\_STRATA STRATA   
ON STRATA.AREA\_ID = AGECOMP.AREA\_ID  
  
-- Filter data results  
WHERE 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)) %>%   
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra()

| **AGE** | **POPULATION\_COUNT** | **SEX** |
| --- | --- | --- |
| 1 | 109,712,069 | 2 |
| 2 | 206,608,351 | 2 |
| 3 | 464,458,137 | 2 |
| 4 | 237,658,858 | 2 |
| 5 | 235,969,541 | 2 |
| 6 | 166,817,093 | 2 |

figure <- ggplot2::ggplot(  
 data = dat0,   
 mapping =   
 aes(x = age,  
 y = population\_count,   
 fill = sex)) +  
 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::ggtitle(label = "EBS Walleye Pollock Age Compositions 1982 – 2022") +   
 ggplot2::guides(fill = guide\_legend(title = "Sex"))+  
 ggplot2::theme\_bw()  
  
figure

|  |
| --- |
| Ex. 3: EBS Walleye Pollock Age Compositions and Age Pyramid. |

### 8.0.9 Ex. NBS Pacific cod biomass and abundance

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

dat <- RODBC::sqlQuery(channel = channel,   
 query =   
 "  
-- Manipulate data to join to  
WITH FILTERED\_STRATA AS (  
SELECT   
AREA\_ID,   
AREA\_NAME,   
DESCRIPTION   
FROM GAP\_PRODUCTS.AKFIN\_AREA  
WHERE AREA\_TYPE in ('STRATUM') AND   
SURVEY\_DEFINITION\_ID = 143)   
  
-- Select columns for output data  
SELECT   
BIOMASS.BIOMASS\_MT,   
BIOMASS.POPULATION\_COUNT,   
BIOMASS.YEAR,   
STRATA.AREA\_NAME  
  
-- Identify what tables to pull data from  
FROM GAP\_PRODUCTS.AKFIN\_BIOMASS BIOMASS   
JOIN FILTERED\_STRATA STRATA   
ON STRATA.AREA\_ID = BIOMASS.AREA\_ID  
  
-- Filter data results  
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)) %>%   
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra() %>%  
 flextable::colformat\_num(x = ., j = "YEAR", big.mark = "")

| **BIOMASS\_MT** | **POPULATION\_COUNT** | **YEAR** | **AREA\_NAME** |
| --- | --- | --- | --- |
| 107,096.730 | 102,734,142 | 2019 | Inner Domain |
| 76,708.433 | 39,605,860 | 2023 | Inner Domain |
| 132,490.152 | 66,187,245 | 2017 | Inner Domain |
| 96,500.697 | 60,433,135 | 2022 | Inner Domain |
| 7,462.559 | 4,724,153 | 2010 | Inner Domain |
| 95,849.983 | 68,767,498 | 2021 | 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

|  |
| --- |
| Ex. 4: NBS Pacific cod biomass and abundance. |

### 8.0.10 Ex. 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 columns for output data  
SELECT   
SURVEY\_DEFINITION\_ID,   
BIOMASS\_MT,   
BIOMASS\_VAR,   
YEAR  
  
-- Identify what tables to pull data from  
FROM GAP\_PRODUCTS.AKFIN\_BIOMASS  
  
-- Filter data results  
WHERE SPECIES\_CODE = 30060   
AND SURVEY\_DEFINITION\_ID = 47   
AND AREA\_ID = 99903   
AND YEAR BETWEEN 1984 AND 2023;") %>%   
 janitor::clean\_names() %>%   
 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)) %>%  
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra() %>%  
 flextable::colformat\_num(x = ., j = "year", big.mark = "")

| **survey\_definition\_id** | **biomass\_mt** | **biomass\_var** | **year** | **biomass\_kmt** | **biomass\_kci\_up** | **biomass\_kci\_dw** |
| --- | --- | --- | --- | --- | --- | --- |
| 47 | 157,295.1 | 2,221,176,968 | 1990 | 157.2951 | 251.5538 | 63.03638 |
| 47 | 483,622.6 | 11,803,384,787 | 1993 | 483.6226 | 700.9093 | 266.33581 |
| 47 | 771,412.8 | 41,434,152,202 | 1996 | 771.4128 | 1,178.5204 | 364.30515 |
| 47 | 727,063.5 | 150,983,542,178 | 1999 | 727.0635 | 1,504.1955 | -50.06854 |
| 47 | 673,155.1 | 49,285,342,922 | 2001 | 673.1551 | 1,117.1611 | 229.14901 |
| 47 | 457,421.6 | 5,186,126,529 | 2003 | 457.4216 | 601.4511 | 313.39204 |

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,   
 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::ggtitle(  
 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

|  |
| --- |
| Ex. 5: GOA Pacific Ocean perch biomass and line plot. |

# 9. Access API data using R

AKFIN has developed web services (apis) to distribute GAP data. Like the GAP\_PRODUCTS schema, these are under active development. These do not require VPN or an oracle connection but they are protected by Oracle authentication, please contact matt.callahan@noaa.gov for information on how to get an api token to use this option.

The url structure is “https://apex.psmfc.org/akfin/data\_marts/gap\_products/gap\_[base table name]” . For example “https://apex.psmfc.org/akfin/data\_marts/gap\_products/gap\_biomass” is the base url to get data from the akfin\_biomass table. Web services linked to large tables have mandatory parameters to reduce data download size. For example to get agecomp data for Bering Sea pollock in area\_id 10 in 2022 you would use “https://apex.psmfc.org/akfin/data\_marts/gap\_products/gap\_biomass?survey\_definition\_id=98&area\_id=10&species\_code=21740&start\_year=2022&end\_year=2022”.

If you’re using R to pull data through web services you might find the [akfingapdata](https://github.com/MattCallahan-NOAA/akfingapdata/tree/main) (pronounced akfin-gap-data not ak-eff-ing-app-data) R package helpful.

# 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()  
}

## 9.1 Ex. Direct database query in R using the (akfingapdata readme)[https://github.com/MattCallahan-NOAA/akfingapdata/blob/main/README.Rmd] R package:

# load packages  
library(odbc)  
library(getPass)  
library(tidyverse)  
  
# connect to AKFIN Oracle database  
con <- dbConnect(odbc::odbc(), "akfin", UID=getPass(msg="USER NAME"), PWD=getPass())

# define species code for pollock  
my\_species <- 21740  
  
#query database  
data<- dbFetch(dbSendQuery(con,  
 paste0("select \* from gap\_products.akfin\_biomass   
where species\_name = ", my\_species,   
" and survey\_definition\_id = 98,   
and area\_id = 10"))) %>%  
rename\_with(tolower) # everyone likes lower case letters better  
  
head(data)

## 9.2 Ex. Direct database query in R using the (akfingapdata readme)[https://github.com/MattCallahan-NOAA/akfingapdata/blob/main/README.Rmd] R package:

library(akfingapdata)  
  
# Sign into akfin with token (need to request token from AKFIN)  
token <- akfingapdata::create\_token(file = paste0(dirname(here::here()), "/akfin\_token.txt"))  
  
akfingapdata::get\_gap\_catch()[,1:6] %>%   
 head() %>%   
 flextable::flextable() %>%  
 flextable::theme\_zebra()

| **cruisejoin** | **hauljoin** | **catchjoin** | **species\_code** | **weight\_kg** | **count** |
| --- | --- | --- | --- | --- | --- |
| -615 | -11,563 | -323,264 | 21,740 | 32.44 | 145 |
| -612 | -13,726 | -377,594 | 21,740 | 39.39 | 30 |
| -612 | -13,728 | -377,650 | 21,740 | 3.22 | 3 |
| -612 | -13,696 | -376,706 | 21,740 | 9.49 | 8 |
| -612 | -13,702 | -376,916 | 21,740 | 18.10 | 21 |
| -615 | -11,533 | -322,193 | 21,740 | 40.81 | 27 |

# 10. 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.

## 10.1 Data tables

### 10.1.1 FOSS\_CATCH

These datasets, FOSS\_CATCH, FOSS\_CPUE\_PRESONLY, FOSS\_HAUL, and FOSS\_SPECIES, when full joined by the HAULJOIN variable, includes zero-filled (presence and absence) observations and catch-per-unit-effort (CPUE) estimates for all identified species at for index stations. 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. For more information about codes used in the tables, please refer to the survey code books (https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual). These data were last updated February 10, 2024.

Number of rows: 928,931

Number of columns: 7

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

COUNT

Taxon count

count, whole number resolution

NUMBER(38,0)

Total whole number of individuals caught in haul or samples collected.

CPUE\_KGKM2

Weight CPUE (kg/km2)

kilograms per kilometers squared

NUMBER(38,6)

Catch weight (kilograms) per unit effort (area swept by the net, units square kilometers).

CPUE\_NOKM2

Number CPUE (no/km2)

count per kilometers squared

NUMBER(38,6)

Numerical catch per unit effort (area swept by the net, units square kilometers).

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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: **‘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)](http://apps-afsc.fisheries.noaa.gov/Publications/ProcRpt/PR2009-04.pdf), [Species identification confidence in the eastern Bering Sea slope survey (1976-2010)](http://apps-afsc.fisheries.noaa.gov/Publications/ProcRpt/PR2014-05.pdf), and [Species identification confidence in the Gulf of Alaska and Aleutian Islands surveys (1980-2011)](http://apps-afsc.fisheries.noaa.gov/Publications/ProcRpt/PR2014-01.pdf).

WEIGHT\_KG

Sample or taxon weight (kg)

kilograms

NUMBER(38,3)

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

### 10.1.2 FOSS\_HAUL

These datasets, FOSS\_CATCH, FOSS\_CPUE\_PRESONLY, FOSS\_HAUL, and FOSS\_SPECIES, when full joined by the HAULJOIN variable, includes zero-filled (presence and absence) observations and catch-per-unit-effort (CPUE) estimates for all identified species at for index stations. 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. For more information about codes used in the tables, please refer to the survey code books (https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual). These data were last updated February 10, 2024.

Number of rows: 32,626

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)

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 integer 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)

Unique integer ID 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. All dates and times are in Alaska time (AKDT) of Anchorage, AK, USA (UTC/GMT -8 hours).

DEPTH\_M

Depth (m)

degrees Celsius

NUMBER(38,1)

Bottom depth (meters).

DISTANCE\_FISHED\_KM

Distance fished (km)

degrees Celsius

NUMBER(38,3)

Distance the net fished (thousands 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)

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

Haul performance code

category

NUMBER(38,0)

This denotes what, if any, issues arose during the haul. For more information, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SRVY

Survey

text abbreviated

VARCHAR2(255 BYTE)

Abbreviated survey names. The column ‘srvy’ is associated with the ‘survey’ and ‘survey\_definition\_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 region. 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\_definition\_id’ columns.

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SURVEY\_NAME

Survey name

text

VARCHAR2(255 BYTE)

Long name of the survey conducted

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

YEAR

Survey year

year

NUMBER(10,0)

Year the observation (survey) was collected.

### 10.1.3 FOSS\_SPECIES

These datasets, FOSS\_CATCH, FOSS\_CPUE\_PRESONLY, FOSS\_HAUL, and FOSS\_SPECIES, when full joined by the HAULJOIN variable, includes zero-filled (presence and absence) observations and catch-per-unit-effort (CPUE) estimates for all identified species at for index stations. 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. For more information about codes used in the tables, please refer to the survey code books (https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual). These data were last updated February 10, 2024.

Number of rows: 1,890

Number of columns: 6

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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/).

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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/).

### 10.1.4 FOSS\_SURVEY\_SPECIES

This reference dataset contains the full list of species by survey to be used to zero-fill FOSS\_CATCH and FOSS\_HAUL for each survey. 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. For more information about codes used in the tables, please refer to the survey code books (https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual). These data were last updated February 10, 2024.

Number of rows: 5,025

Number of columns: 2

Column name from data

Descriptive column Name

Units

Oracle data type

Column description

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

SURVEY\_DEFINITION\_ID

Survey ID

ID code

NUMBER(38,0)

The survey definition ID code uniquely identifies a survey/survey design. The column ‘survey\_definition\_id’ is associated with the ‘srvy’ and ‘survey’ columns. For a complete list of surveys, review the [code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

### 10.1.5 FOSS\_TAXON\_GROUP

This reference dataset contains suggested search groups for simplifying species selection in the FOSS data platform so users can better search through FOSS\_CATCH. 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. For more information about codes used in the tables, please refer to the survey code books (https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual). These data were last updated February 10, 2024.

Number of rows: 33,721

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

Taxonomic rank

category

VARCHAR2(255 BYTE)

The taxonomic rank of a taxon identification.

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](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

# 11. Using the FOSS platform

|  |
| --- |
| AFSC Groundfish and Crab Assessment Program Bottom Trawl Survey data interface on the Fisheries One Stop Shop platform. |

## 11.1 Select and filter

Select, filter, and package this and other NOAA Fisheries data from the [Fisheries One Stop Shop (FOSS)](https://www.fisheries.noaa.gov/foss) platform. A user guide for the FOSS platform can be found [here](https://www.fisheries.noaa.gov/foss/f?p=215:7:7542600605674:::::). 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 Speices 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.

|  |
| --- |
| Diagram of selection and search tools available on the FOSS platfrom. |

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](https://www.fisheries.noaa.gov/alaska/science-data/groundfish-assessment-program-bottom-trawl-surveys#data-products).
* Year: Surveys are not conducted in all years, so only data from the years for which the survey was conducted will be returned.
* Species: Common name of all species ever encountered in the survey. Find more information about these species in our [survey code books](https://www.fisheries.noaa.gov/resource/document/groundfish-survey-species-code-manual-and-data-codes-manual).

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.

|  |
| --- |
| Diagram of selection and search tools available on the FOSS platofrom. |

## 11.2 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](https://github.com/afsc-gap-products/akgfmaps)), 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 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).

|  |
| --- |
| Diagram of the pre-set data format options. |

## 11.3 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.

|  |
| --- |
| Example data returned from running the report. |

# 12. Access via API and 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](https://docs.oracle.com/en/database/oracle/oracle-rest-data-services/22.3/books.html#AELIG90103/). Useful introductions to using APIs in R can be found [here](https://www.dataquest.io/blog/r-api-tutorial/).

## 12.1 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\_survey\_id"), big.mark = "")

| **year** | **srvy** | **survey** | **survey\_id** | **cruise** | **haul** | **stratum** | **station** | **vessel\_name** | **vessel\_id** | **date\_time** | **latitude\_dd** | **longitude\_dd** | **species\_code** | **common\_name** | **scientific\_name** | **taxon\_confidence** | **cpue\_kgha** | **cpue\_kgkm2** | **cpue\_kg1000km2** | **cpue\_noha** | **cpue\_nokm2** | **cpue\_no1000km2** | **weight\_kg** | **count** | **bottom\_temperature\_c** | **surface\_temperature\_c** | **depth\_m** | **distance\_fished\_km** | **net\_width\_m** | **net\_height\_m** | **area\_swept\_ha** | **duration\_hr** | **tsn** | **ak\_survey\_id** | **links** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2002 | AI | Aleutian Islands Bottom Trawl Survey | 52 | 200201 | 6 | 722 | 307-63 | Vesteraalen | 94 | 05/17/2002 18:56:58 | 53.737 | -167.016 | 95020 | feathery bryozoan | Eucratea loricata | Low | 0.017 | 1.749 | 1,749.445 |  |  |  | 0.044 | 0 | 4.1 | 5.3 | 187 | 1.561 | 16.112 | 7.25 | 2.515 | 0.28 | 155809 | 1917453 | [[data.frame]] |
| 2002 | AI | Aleutian Islands Bottom Trawl Survey | 52 | 200201 | 6 | 722 | 307-63 | Vesteraalen | 94 | 05/17/2002 18:56:58 | 53.737 | -167.016 | 79000 | squid unid. | Decapodiformes | High | 0.022 | 2.227 | 2,226.567 | 3.181 | 318.081 | 318,080.93 | 0.056 | 8 | 4.1 | 5.3 | 187 | 1.561 | 16.112 | 7.25 | 2.515 | 0.28 |  | 1917454 | [[data.frame]] |
| 2002 | AI | Aleutian Islands Bottom Trawl Survey | 52 | 200201 | 6 | 722 | 307-63 | Vesteraalen | 94 | 05/17/2002 18:56:58 | 53.737 | -167.016 | 24191 | shortfin eelpout | Lycodes brevipes | High | 0.036 | 3.578 | 3,578.410 | 0.795 | 79.520 | 79,520.23 | 0.090 | 2 | 4.1 | 5.3 | 187 | 1.561 | 16.112 | 7.25 | 2.515 | 0.28 | 165258 | 1917455 | [[data.frame]] |

## 12.2 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"

## 12.3 Ex. 3: Filter by Year

Show all the data greater than the year 2020.

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::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra() %>%  
 flextable::colformat\_num(x = ., j = c("year", "species\_code"), big.mark = "")

| **year** | **srvy** | **stratum** | **species\_code** | **cpue\_kgkm2** |
| --- | --- | --- | --- | --- |
| 2022 | AI | 793 | 80540 | 0.361 |
| 2022 | AI | 793 | 401 | 0.903 |
| 2022 | AI | 793 | 20006 | 1.661 |

## 12.4 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)) %>%  
 dplyr::select(year, srvy, stratum, species\_code, cpue\_kgkm2) %>%  
 flextable::flextable() %>%  
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra() %>%  
 flextable::colformat\_num(x = ., j = c("year", "species\_code"), big.mark = "")

| **year** | **srvy** | **stratum** | **species\_code** | **cpue\_kgkm2** |
| --- | --- | --- | --- | --- |
| 2002 | AI | 722 | 21740 | 775.322 |
| 2002 | AI | 722 | 21740 | 10,685.806 |
| 2002 | AI | 721 | 21740 | 0.640 |

## 12.5 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":"%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)) %>%  
 dplyr::select(year, srvy, stratum, species\_code, cpue\_kgkm2) %>%  
 flextable::flextable() %>%  
 flextable::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra() %>%  
 flextable::colformat\_num(x = ., j = c("year", "species\_code"), big.mark = "")

| **year** | **srvy** | **stratum** | **species\_code** | **cpue\_kgkm2** |
| --- | --- | --- | --- | --- |
| 2022 | AI | 793 | 21740 | 7,853.632 |
| 2022 | AI | 721 | 21740 | 7,235.010 |
| 2022 | AI | 722 | 21740 | 22,754.334 |

## 12.6 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::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra() %>%  
 flextable::colformat\_num(x = ., j = c("year", "species\_code"), big.mark = "")

| **year** | **srvy** | **stratum** | **species\_code** | **cpue\_kgkm2** |
| --- | --- | --- | --- | --- |
| 1989 | EBS | 10 | 66548 | 1.164 |
| 1989 | EBS | 10 | 69322 | 1.164 |
| 1989 | EBS | 10 | 43000 | 2.353 |

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

Pacific cod catch-per-unit-effort estimates for NBS in 2021 and map constructed using [akgfmaps](https://github.com/afsc-gap-products/akgfmaps).

# 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::fit\_to\_width(max\_width = 6) %>%   
 flextable::theme\_zebra()

| **stratum** | **station** | **latitude\_dd** | **longitude\_dd** | **cpue\_kgkm2** |
| --- | --- | --- | --- | --- |
| 81 |  | 61.66434 | -172.2655 | 2,895.258 |
| 81 |  | 62.33740 | -173.1702 | 1,235.545 |
| 70 |  | 62.03713 | -171.6528 | 0.000 |

# devtools::install\_github("afsc-gap-products/akgfmaps", build\_vignettes = TRUE)  
library(akgfmaps)  
  
figure <- akgfmaps::make\_idw\_map(  
 CPUE\_KGHA = data$cpue\_kgkm2, # calculates the same, regardless of units.   
 LATITUDE = data$latitude\_dd,   
 LONGITUDE = data$longitude\_dd,   
 region = "bs.north", # Predefined EBS area  
 set.breaks = "jenks", # Gets Jenks breaks from classint::classIntervals()  
 in.crs = "+proj=longlat", # Set input coordinate reference system  
 out.crs = "EPSG:3338", # Set output coordinate reference system  
 grid.cell = c(20000, 20000))$plot + # 20x20km grid  
 ggplot2::guides(fill=guide\_legend(title = "Pacific cod\nCPUE (kg/km2)"))

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

figure

|  |
| --- |
| Ex. 7: Visualize CPUE data in distribution map. |

# 13. Access via API and Python

### 13.0.1 {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 Python and R.   
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](https://pyafscgap.org).

### 13.0.2 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](https://www.oracle.com/database/technologies/appdev/rest.html) 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()

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

### 13.0.3 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.

### 13.0.4 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.

### 13.0.5 Load into Pandas

The results from the afscgap package are serializable and can be loaded into other tools like [Pandas](https://pandas.pydata.org/). 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')  
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.

### 13.0.6 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](https://www.oracle.com/database/technologies/appdev/rest.html) query generation which dictates how the library communicates with the API service. However, users can provide raw ORDS queries as well using [manual filtering](https://pyafscgap.org/devdocs/afscgap.html#manual-filtering).

### 13.0.7 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.

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 set\_presence\_only(True) | Count with set\_presence\_only(False) |
| --- | --- | --- |
| 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.

### 13.0.8 More information

Please see the [API documentation](https://pyafscgap.org/devdocs/afscgap.html) for the Python library for additional details.

# 14. Access via Oracle and 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 GAP\_PRODUCTS Oracle schema.

### 14.0.1 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.

library(gapindex)  
channel <- gapindex::get\_connected()

### 14.0.2 Ex. 1: Join data

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

SELECT   
hh.YEAR,  
hh.SRVY,   
hh.SURVEY,  
hh.SURVEY\_DEFINITION\_ID,  
hh.SURVEY\_NAME,  
hh.CRUISE,  
hh.CRUISEJOIN,   
hh.HAUL,  
hh.HAULJOIN,  
hh.STRATUM,  
hh.STATION,  
hh.VESSEL\_ID,  
hh.VESSEL\_NAME,   
hh.DATE\_TIME,  
hh.LATITUDE\_DD\_START,   
hh.LONGITUDE\_DD\_START,   
hh.LATITUDE\_DD\_END,  
hh.LONGITUDE\_DD\_END,   
hh.BOTTOM\_TEMPERATURE\_C,  
hh.SURFACE\_TEMPERATURE\_C,  
hh.DEPTH\_M,  
cc.SPECIES\_CODE,  
ss.ITIS,  
ss.WORMS,  
ss.COMMON\_NAME,   
ss.SCIENTIFIC\_NAME,  
ss.ID\_RANK,  
CASE WHEN cc.CPUE\_KGKM2 IS NULL THEN 0 ELSE cc.CPUE\_KGKM2 END AS CPUE\_KGKM2,  
CASE WHEN cc.CPUE\_NOKM2 IS NULL THEN 0 ELSE cc.CPUE\_NOKM2 END AS CPUE\_NOKM2,  
CASE WHEN cc.COUNT IS NULL THEN 0 ELSE cc.COUNT END AS COUNT,  
CASE WHEN cc.WEIGHT\_KG IS NULL THEN 0 ELSE cc.WEIGHT\_KG END AS WEIGHT\_KG,  
CASE WHEN cc.TAXON\_CONFIDENCE IS NULL THEN NULL ELSE cc.TAXON\_CONFIDENCE END AS TAXON\_CONFIDENCE,  
hh.AREA\_SWEPT\_KM2,   
hh.DISTANCE\_FISHED\_KM,  
hh.DURATION\_HR,   
hh.NET\_WIDTH\_M,  
hh.NET\_HEIGHT\_M,  
hh.PERFORMANCE   
FROM GAP\_PRODUCTS.FOSS\_SURVEY\_SPECIES sv  
FULL OUTER JOIN GAP\_PRODUCTS.FOSS\_SPECIES ss  
ON sv.SPECIES\_CODE = ss.SPECIES\_CODE  
FULL OUTER JOIN GAP\_PRODUCTS.FOSS\_HAUL hh  
ON sv.SURVEY\_DEFINITION\_ID = hh.SURVEY\_DEFINITION\_ID  
FULL OUTER JOIN GAP\_PRODUCTS.FOSS\_CATCH cc  
ON sv.SPECIES\_CODE = cc.SPECIES\_CODE  
AND hh.HAULJOIN = cc.HAULJOIN

### 14.0.3 Ex. 2: Subset data

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 SPECIES\_CODE = 21720 -- 'Pacific cod'   
AND YEAR >= 2010   
AND YEAR < 2021")  
  
head(a)

HAULJOIN SPECIES\_CODE CPUE\_KGKM2 CPUE\_NOKM2 COUNT WEIGHT\_KG TAXON\_CONFIDENCE  
1 -19288 21720 449.83013 1876.17592 83 19.90 1  
2 -19252 21720 413.48285 248.08971 12 20.00 1  
3 -17850 21720 990.13573 152.26165 7 45.52 1  
4 -18165 21720 1053.27228 241.85357 12 52.26 1  
5 -18731 21720 946.34812 2592.13274 118 43.08 1  
6 -16960 21720 89.15857 19.99071 1 4.46 1  
 YEAR SRVY SURVEY SURVEY\_DEFINITION\_ID  
1 2019 EBS eastern Bering Sea 98  
2 2019 EBS eastern Bering Sea 98  
3 2018 EBS eastern Bering Sea 98  
4 2018 EBS eastern Bering Sea 98  
5 2019 EBS eastern Bering Sea 98  
6 2017 EBS eastern Bering Sea 98  
 SURVEY\_NAME CRUISE CRUISEJOIN  
1 Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey 201901 -727  
2 Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey 201901 -727  
3 Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey 201801 -723  
4 Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey 201801 -723  
5 Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey 201901 -726  
6 Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey 201701 -712  
 HAULJOIN.1 HAUL STRATUM STATION VESSEL\_ID VESSEL\_NAME DATE\_TIME  
1 -19288 96 20 O-18 162 ALASKA KNIGHT 2019-06-29 06:54:00  
2 -19252 76 31 G-03 162 ALASKA KNIGHT 2019-06-24 15:52:02  
3 -17850 117 32 F-19 162 ALASKA KNIGHT 2018-07-02 09:49:43  
4 -18165 172 62 Q-27 162 ALASKA KNIGHT 2018-07-26 07:49:26  
5 -18731 11 31 I-13 94 VESTERAALEN 2019-06-04 13:15:57  
6 -16960 155 50 E-21 94 VESTERAALEN 2017-07-17 15:16:56  
 LATITUDE\_DD\_START LONGITUDE\_DD\_START LATITUDE\_DD\_END LONGITUDE\_DD\_END  
1 59.68079 -168.6144 59.65546 -168.6178  
2 57.01591 -166.4752 56.99137 -166.4601  
3 56.67170 -168.9406 56.67359 -168.8919  
4 60.31173 -174.7032 60.33716 -174.7090  
5 57.69052 -160.2580 57.66518 -160.2640  
6 56.34430 -170.0621 56.32694 -170.0948  
 BOTTOM\_TEMPERATURE\_C SURFACE\_TEMPERATURE\_C DEPTH\_M DISTANCE\_FISHED\_KM  
1 5.1 7.8 39 2.821  
2 4.1 9.7 74 2.880  
3 4.5 8.8 99 3.005  
4 3.2 10.1 103 2.845  
5 5.5 7.5 54 2.840  
6 3.9 7.9 110 2.796  
 DURATION\_HR NET\_WIDTH\_M NET\_HEIGHT\_M AREA\_SWEPT\_KM2 PERFORMANCE  
1 0.505 15.682 2.227 0.044239 0  
2 0.528 16.795 2.126 0.048370 0  
3 0.525 15.299 2.152 0.045973 0  
4 0.511 17.440 2.200 0.049617 0  
5 0.520 16.029 2.200 0.045522 0  
6 0.508 17.891 1.971 0.050023 0

### 14.0.4 Ex. 3: Find all species found in the eastern Bering Sea (EBS) survey in 2023

# Pull data  
a <- RODBC::sqlQuery(  
channel = channel,   
query =   
"SELECT DISTINCT   
ss.COMMON\_NAME,  
ss.SCIENTIFIC\_NAME,   
ss.ID\_RANK,   
ss.WORMS  
FROM GAP\_PRODUCTS.FOSS\_CATCH cc -- get species codes  
LEFT JOIN GAP\_PRODUCTS.FOSS\_SPECIES ss -- get species info  
ON cc.SPECIES\_CODE = ss.SPECIES\_CODE  
LEFT JOIN GAP\_PRODUCTS.FOSS\_HAUL hh -- filter by year and survey  
ON cc.HAULJOIN = hh.HAULJOIN  
WHERE hh.YEAR = 2023  
AND hh.SURVEY\_DEFINITION\_ID = 98 -- EBS survey  
ORDER BY COMMON\_NAME")  
  
head(a)

COMMON\_NAME SCIENTIFIC\_NAME ID\_RANK WORMS  
1 Alaska great-tellin Megangulus luteus species 423511  
2 Alaska plaice Pleuronectes quadrituberculatus species 254564  
3 Alaska razor Siliqua alta species 413689  
4 Alaska skate Arctoraja parmifera species 1577324  
5 Alaska skate egg case Arctoraja parmifera egg case species NA  
6 Alaskan hermit Pagurus ochotensis species 366742

# 15. Open source code

## 15.1 R Packages

### 15.1.1 [akgfmaps R package](https://github.com/afsc-gap-products/akgfmaps)

Bttom trawl survey maps layers and plotting examples. **POC:** Sean Rohan

### 15.1.2 [coldpool R package](https://github.com/afsc-gap-products/coldpool)

Cold pool area and temperature data products for the Bering Sea. **POC:** Sean Rohan

### 15.1.3 [akfishcondition R package](https://github.com/afsc-gap-products/akfishcondition)

Groundfish morphometric condition indicators for fish in the Bering Sea, Aleutian Islands, and Gulf of Alaska. **POC:** Sean Rohan

### 15.1.4 [gapindex R package](https://github.com/afsc-gap-products/gapindex)

Calculation of Design-Based Indices of Abundance and Composition for AFSC GAP Bottom Trawl Surveys. **POC:** Zack Oyafuso and Margaret Siple

# 16. Production run notes

# 17. R Version Metadata

R version 4.3.1 (2023-06-16 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\_4.3.1 fastmap\_1.1.1 cli\_3.6.1 tools\_4.3.1   
 [5] htmltools\_0.5.7 rstudioapi\_0.15.0 yaml\_2.3.8 rmarkdown\_2.25   
 [9] knitr\_1.45 jsonlite\_1.8.8 xfun\_0.41 digest\_0.6.33   
[13] rlang\_1.1.2 evaluate\_0.23

### 17.0.1 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 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.

### 17.0.2 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.

# 18. Acknowledgments

# 19. 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.

# 20. Land Acknowledgements

We would like to thank the many communities of the Bering Strait region and their members who have helped contribute to this document. The knowledge, experiences, and insights of the people of the Bering Strait region 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, including the local names used for the species covered by this document, identifying species of interest or concern that should be included in this document, and participation in an open dialog about how we can improve our collective knowledge of the ecosystem and the region.

NOAA Fisheries Alaska Fisheries Science Center’s work is conducted in the waters and along the coastlines of Alaska, which include the traditional home lands and waters of the Inupiat, Yupiit, Siberian Yupiit, Unangax, Alutiiq/Sugpiaq, Eyak, Dena’ina Athabascan, Tlingit, Haida, and Tsimshian who have stewarded their lands and waters since time immemorial. We are indebted to these peoples for their wisdom and knowledge of their lands and waters.

This document was prepared in the greater Seattle area, which are the traditional lands of the Coast Salish people, including the Duwamish people, past and present. We are grateful for their continued sharing of vision, wisdom, values, and leadership.

# 21. Technical Acknowledgments

This quarto book is based off the [NOAA-quarto-book](https://github.com/nmfs-opensci/NOAA-quarto-book) GitHub repo designed by Eli Holmes.

This repo and GitHub Action was based on the tutorial by Openscapes [quarto-website-tutorial](https://github.com/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.

## 21.2 Collaborators

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

* [Alaska Stock Assessments](https://www.fisheries.noaa.gov/alaska/population-assessments/alaska-stock-assessments)
* [North Pacific Groundfish Stock Assessment and Fishery Evaluation Reports](https://www.fisheries.noaa.gov/alaska/population-assessments/north-pacific-groundfish-stock-assessment-and-fishery-evaluation)
* [Groundfish Economic Status Reports for the Gulf of Alaska and Bering Sea and Aleutian Islands](https://www.fisheries.noaa.gov/alaska/commercial-fishing/groundfish-economic-status-reports-gulf-alaska-and-bering-sea-and-aleutian-islands)
* [Alaska Marine Ecosystem Status Report Database](https://www.fisheries.noaa.gov/resource/data/alaska-marine-ecosystem-status-report-archive)
* [Southeast Alaska Coastal Monitoring Survey Reports](https://www.fisheries.noaa.gov/alaska/commercial-fishing/southeast-alaska-coastal-monitoring-survey-reports)
* [Alaska Fisheries Life History Database](https://www.fisheries.noaa.gov/resource/data/alaska-fisheries-life-history-database)
* [Essential Fish Habitat Research Plan in Alaska](https://www.fisheries.noaa.gov/alaska/habitat-conservation/essential-fish-habitat-research-plan-alaska)

# 22. References

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.

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., Raring, N. W., Siple, M. C., Dowlin, A. N., Riggle, B. C., and Laman, E. A. and. (2023). *Data report: 2022 Aleutian Islands bottom trawl survey* (AFSC Processed Rep. No. 2023-07; p. 230). U.S. Dep. Commer. <https://doi.org/10.25923/85cy-g225>