**DATA BIOGRAPHY**

Your Name: Alana Santana, Rory Spurr

**1. RESOURCE OVERVIEW** (*what, why, who*)

1.1. Title of the dataset. *Descriptive title that usually includes data type, time period, location, and name of author, program, or institution.*

NOAA West Coast Region (WCR) Permitting Data, 2012 to 2023, U.S. West Coast (WA, OR, CA, ID), export from larger relational database maintained by NOAA WCR (Protected Resources Division).

1.2. Abstract. *One paragraph summary of the dataset in plain language. Include one sentence of broader context, followed by the dataset’s origin/purpose, and briefly expand on the elements of the title. A length of 200-250 words is a good target.*

The NOAA WCR permitting database stores information about permits they issue for conducting research activities that may harm (purposefully or incidentally) species listed as threatened or endangered under the Endangered Species Act (ESA). The database collates information submitted by permit applicants through NOAA's online application system for Authorizations and Permits for Protected Species (APPS; <https://apps.nmfs.noaa.gov/>). The database is used by the Protected Resources Division (PRD) to store information on the amount of take being authorized for threatened and endangered species. This in turn allows the PRD to authorize permits that minimize harm to the species while maximizing the conservation benefit gained from the project. For example, the database provides information on the permit type, whether the permit was issued, where the research is occurring, what type of research will be conducted, what type of ESA-listed species are involved, and the amount of take authorized during the period of the project. Permit types can range from streamlined projects conducted on threatened species (4d and tribal 4d) to permits that allow research to be conducted on endangered species (10a1A permits). The portion of the dataset that the UW capstone team is working with is a subset that was queried from the larger relational database, and includes only permit types specific to scientific research actions.

1.3. Purpose. *Brief description of why the data were collected, including the goals and intended outcomes (this may or may not include application to decision-making).*

The NOAA WCR permitting dataset was collected as a form of record-keeping of current and past permits. The data is automatically stored in this repository to be reaccessed by anyone searching for past or current permits. The data is important for making decisions on issuing future permits, as NOAA can view the amount of take being authorized on an ESA-listed species, and determine if the amount of take being requested will put the species in jeopardy. Along with this, NOAA can review the methods and goals of the proposed study, and decide if it has a bona fide and desirable purpose in enhancing the survivability of the species.

1.4. Contacts. *Provide contact info for the people who managed the project, collected the data, generated the dataset, and/or managed the data. Contact information should include name, organization, role in the project, email and/or phone.*

Agency: NOAA

Region: West Coast Region

Office Phone: (503) 230-5400

Website: <https://www.fisheries.noaa.gov/about/west-coast-region>

Address: 1201 Northeast Lloyd Boulevard, Suite 1100, Portland OR 97232

1.5. Sponsors. *Who or what organization sponsored collection of the data (e.g., NOAA as a part of a mandated monitoring program)? Who funded collection of the data (if applicable)?*

NOAA is the federal agency overseeing the Protected Resource Division through the Department of Commerce. Under the Protected Resource Division, the Regional Permit Coordinator manages and allocates scientific research and take permits. All data regarding permits applications, issuance, withdrawal, or denial is recorded and stored within a database. NOAA is the sponsoring organization for the collection of this data. As a federally mandated program, the collection of data is federally funded.

1.6. Citation for the dataset. *Use the citation format below and include a link to the data source.*

National Marine Fisheries Service (NMFS) and Oregon Department of Fish and Wildlife (ODFW). Authorizations and Permits for Protected Species (APPS). Current authorizations for research under ESA Section 10(a)(1)(A) and Section 4(d), and reported take from 2012-2022, for fish species in Washington, Oregon, Idaho, and California. Available online at https://apps.nmfs.noaa.gov/. Accessed 02/09/23.

1.7. Keywords. *Include 3-5 keywords for the dataset. Think of these as search terms that someone might use to find the data.*

Research Permits, West Coast Region, NOAA Fisheries, ESA

**2. TEMPORAL AND SPATIAL EXTENTS** (*when, where*)

2.1. Temporal extent. *The entire time range (specific years) for observations included in the dataset.*

The dataset includes permitting data from October of 2010 to the present. However, the data made available for this project extends to November of 2022.

2.2. Temporal resolution. *The frequency at which data are collected or acquired. Be as specific as possible. Note whether measurements were taken at regular intervals or irregularly.*

The APPS database is continuously being updated and accessed by permit managers.

However, the data we were provided shows a snapshot in time of the data that is ongoing in the APPS database. For each permit inquiry, the date was logged for the request, issuance, withdrawal, and/or denial.

Prior to the data export, permits are reviewed thoroughly at the time of application. However, the permit information is not reviewed unless there is an incident report or the annual report is submitted.

2.3. Spatial extent. *Boundaries of the data set. If possible, include both the (a) geographic description, and (b) coordinates describing north, south, east, and west boundaries of the area included in the data. You do not need to include granular geospatial data (e.g., survey tracks, buoy locations).*

The dataset includes permitting data from the U.S. west coast region, which includes anadromous and marine waters of California, Oregon, Washington, and Idaho. Below is a map regarding the extent of WCR and where the data on permits was collected.



*Fig. 1 Spatial extent of NOAA’s West Coast Region*

2.4. Spatial resolution. Specificity with which spatial data are recorded. For example, are locations of measurements recorded using GPS? Locality (place) names? Were measurements collected on a uniform grid and if so, at what spatial scale?

Spatial resolutions vary between permits, with most of the permits containing a user inputted 8-digit Hydrologic Unit Code (HUC 8) describing where the project will be conducted. For some projects no HUC 8 code was provided (usually in projects encompassing a wide range of HUC 8’s, or projects that are conducted in marine waters) the spatial resolution is larger than a HUC 8, and can generally be described by the waterbody or basin name, or the geographical range occupied by the Evolutionary Significant Unit (ESU) that the research is being conducted on.

**3. RESOURCE CONTENT** (*what*)

3.1. Digital context. *Names of data file(s), names of tables within data file(s), file format(s), and date the data were last modified. If you have multiple data files, describe any relationships among them (e.g., queried database tables saved as separate files and linked through an identifier?). For each data file and table within a data file, include a brief (1 sentence) narrative description of the contents.*

Data is queried by the data manager through the APPS interface.

Data is downloaded as a “.csv” file from the APPS interface.

The following SQL queries are used to download data as .csv files from the APPS database:

Authorized permit data: WCRpermitBiOp\_allregns\_all\_years

SELECT DISTINCT

pac.FileNumber,

pac.ResultCode,

pac.AccountStatus,

pac.PermitStatus,

pac.DateIssued,

pac.DateExpired,

pac.AnnualTimeStart,

pac.AnnualTimeEnd,

pac.ProjectTitle,

pac.FirstName,

pac.LastName,

pac.Organization,

loc.HUCNumber,

loc.WaterbodyName,

loc.BasinName,

loc.StreamName,

loc.LocationDescription,

tak.CommonName,

tak.Population,

tak.Run,

tak.Production,

tak.LifeStage,

tak.Sex,

tak.TakeAction,

tak.CaptureMethod,

tak.ExpTake,

tak.IndMort

FROM dbo.PacificSection10a1AAnd4d\_vw AS pac

INNER JOIN dbo.Location\_vw AS loc ON pac.ProjectID = loc.ProjectID

INNER JOIN dbo.TakeInformation\_vw AS tak ON loc.LocationID = tak.LocationID

WHERE tak.SpeciesID IN (1,2,3,4,5,44,241,243,254,1028)

AND pac.RegionID IN ('WA','OR','ID','CA')

ORDER BY pac.FileNumber

Reported permit data: WCRPermitBiOp\_Pass report data 4d and S10

SELECT DISTINCT

rpt.FileNumber,

rpt.ReportID,

rpt.ResultCode,

rpt.AnnualTimeStart,

rpt.AnnualTimeEnd,

rpt.DateReportPeriodEnd,

rpt.RegionID,

rpt.FirstName,

rpt.LastName,

rpt.Organization,

rpt.ProjectTitle,

lv.BasinName,

lv.WaterbodyName,

lv.StreamName,

tv.CommonName,

tv.Population,

tv.Production,

tv.LifeStage,

tv.TakeAction,

tv.CaptureMethod,

tv.ExpTake,

tv.ActTake,

tv.IndMort,

tv.ActMort

FROM dbo.RptPacSection10a1AAnd4d\_vw AS rpt

INNER JOIN dbo.RptLocation\_vw AS lv ON rpt.ReportID = lv.ReportID

INNER JOIN dbo.RptTake\_vw AS tv ON lv.LocationID = tv.LocationID

WHERE rpt.AnnualTimeStart > '12/31/2010'

AND rpt.RegionID IN ('OR','WA','ID','CA')

AND tv.SpeciesID IN (1,2,3,4,5,44,241,243,254,1028)

ORDER BY rpt.FileNumber, rpt.AnnualTimeStart

Abundance data are compiled from various sources including NOAA Fisheries’ Northwest and Southwest Fisheries Science Centers, state hatchery management programs, and other fecundity, survival, and outmigration estimates assembled by NMFS. A file of compiled abundance data current as of March 2022 was provided by NMFS as Abundance\_2022-03-17.xlsx. It is generated by an R script titled Compile\_abundance.R saved within the permit team’s Section 10 biological opinion github repository.

3.2. Data components and data table attributes. *This section details the contents of each data table and/or data file and might be most effectively organized as a table (but it’s up to you). For each data file/data table, provide the names, definitions, and units of the attributes of any data in tabular format (e.g., column headers in a CSV file). Depending on the nature of the data, this could include: parameter name, measurement units, instrument type, precision, accuracy, taxonomic details, definitions of codes used, and any other important information for an analyst (e.g., quality review notes, missing values). Indicate whether data are raw values (not modified in any way after collection), processed values (corrected or calibrated), or derived values (an index or summarized value calculated based on other data).*

***WCRpermitBiOp\_allregns\_all\_years\_\_7Jan2022.csv***

*Note - data components that are italicized and bolded are unique to WCRPermitBiOp\_Pass report data 4d and S10\_22March22.csv data file.*

| **Data Component** | **Attributes** |
| --- | --- |
| FileNumber | File number automatically assigned by the APPS system; used in correspondence about the application |
| ResultCode | Permit type or authority |
| AccountStatus | Indicates the current status of the Application |
| PermitStatus | Indicates the status of the permit that was applied for. (ex “Issued”, “In Review”) |
| DateIssued | Month, day, and year of the permit issuance. (yyyy-mm-dd) |
| DateExpired | Month, day, and year of the permit expiry. (yyyy-mm-dd) |
| AnnualTimeStart | Window when applicants begin to send in reports. (yyyy-mm-dd) |
| AnnualTimeEnd | Window when applicants cease to send in reports. (yyyy-mm-dd) |
| ProjectTitle | Name of the research project being conducted. |
| FirstName | First name of the principal investigator. |
| LastName | Last name of the principal investigator. |
| Organization | Name of the organization sponsoring the research. |
| HUCNumber | Displays the Hydrologic Unit Code (HUC) where that take is expected to occur. |
| WaterbodyName | Name of the major waterbody where the take is expected to occur. |
| BasinName | Name of the water basin where the take is expected to occur. |
| StreamName | Name or names of the streams involved where the take is expected to occur. Can be general (“Lower Columbia River and its tributaries”) or specific (“Battle Creek”). |
| LocationDescription | Specific description giving location of sampling site/trap etc. where take is expected to occur. Can be general if take is expected to occur all along the stream/waterbody. |
| CommonName | Common name of the species on which take is expected to occur. |
| Population | Specific population on which take is expected to occur. Mainly defined by the river or geographic area. |
| Run | Specific run where take is expected to occur. Usually described as the time of year (Spring, Summer, Winter) when the salmon or other anadromous fish return to spawn. |
| Production | Indicates whether the fish are for hatchery production or wild. |
| LifeStage | Life stage of the species/individual the project expects to take. |
| Sex | Sex of the species/individual the project expects to take. |
| TakeAction | Description of the kind of take that is expected to occur (e.g., “Broodstock collection”, “Capture”, “Handle”). |
| CaptureMethod | Gear used to capture species. |
| ExpTake | Number of individuals the project expects to take as a result of research. |
| IndMort | Number of incidental mortality as a result of research. |

***WCRPermitBiOp\_Pass report data 4d and S10\_22March22.csv***

| **Data Component** | **Attributes** |
| --- | --- |
| FileNumber | File number automatically assigned by the APPS system; used in correspondence about the application. |
| ***ReportID*** | ***Five-digit unique code for each active project that reported take and mortality.*** |
| ResultCode | Permit type or authority. |
| AnnualTimeStart | Window when applicants begin to send in reports. |
| AnnualTimeEnd | Window when applicants cease to send in reports. |
| ***DateReportPeriodEnd*** | ***Date in which the report on actual take and mortality is submitted/ends. Reported in YYYY - MM - DD format.*** |
| ***RegionID*** | ***State abbreviations for where permits are occurring. I.e. WA, CA, OR, ID.*** |
| FirstName | First name of the principal investigator. |
| LastName | Last name of the principal investigator. |
| Organization | Name of the organization sponsoring the research. |
| ProjectTitle | Name of the research project being conducted. |
| BasinName | Name of the water basin where the take is expected to occur |
| WaterbodyName | Name of the major waterbody where the take is expected to occur |
| StreamName | Name or names of the streams involved where the take is expected to occur. Can be general (“Lower Columbia River and its tributaries”) or specific (“Battle Creek”). |
| CommonName | Common name of the species on which take is expected to occur. |
| Population | Specific population on which take is expected to occur.. Mainly defined by the river or geographic area. |
| Production | Indicates whether the fish are for hatchery production or wild. |
| LifeStage | Life stage of the species/individual the project expects to take. |
| TakeAction | Description of the kind of take that is expected to occur (ex. “Broodstock collection”, “Capture”, “Handle”) |
| CaptureMethod | Gear used to capture species. |
| ExpTake | Predicted number of individuals the project expects to take as a result of research. |
| ***ActTake*** | ***Actual number of individuals a project takes that occurred as a result of research.*** |
| IndMort | Predicted number of incidental mortality as a result of research. |
| ***ActMort*** | ***Actual number of lethal take that occurred as a result of research.*** |

***APPS\_HUCassignments\_11Feb23.csv***

| **Data Component** | **Attributes** |
| --- | --- |
| speciesid | Unique code given to each species |
| populationid | Unique code given to each unique ESU/DPS of a species |
| Species | Common name of species |
| Population/Stock | Specific ESU/DPS of species |
| Status | ESA-listing status |
| HUC8 | HUC 8 where species could be found |
| Basin Name | Name of basin containing the HUC 8 |

**4. METHODS** (*how*)

4.1. Lineage statement. *Provide a summary of the methods used to collect the data. Ideally, this is a brief narrative description that includes citations to standard operating procedures, field manuals, or other references.*

Data are collected and recorded when applications are submitted through the Authorizations and Permits for Protected Species (APPS) online application system. When researchers, hatchery managers, or other professionals wish to complete research on an ESA-listed species or plan on encountering ESA-listed species, they submit their applications through the APPS online portal. Data on the type of species the research will be conducted on, the type of take occurring, where the take will be occurring, and how much lethal and non-lethal take is expected are then reviewed by the permits team and either approved, edited, or denied. For permits that are issued, if changes are made to permit information while it is active any approved changes to authorized take are immediately reflected in data exported from APPS.

After the end of each sampling season researchers also submit reporting information on what actually occurred in the field. This includes confirming where work was actually conducted, what gear was used, and recording how many fish of each type (species, life stage, and origin) were actually encountered, sampled, and killed. Permit team staff also review annual reports and correct any errors discovered during review. Any changes or corrections to reporting data are also immediately reflected in data exported from APPS.

4.2. Process steps. *The general process steps that occurred between data collection and its current form (brief narrative description or bulleted list). Depending on the dataset, processing might include digitization, removing or identifying outliers via computer scripts, file processing, data summarization, or data transformations. This does not need to be exhaustive, but should include information that would be important for an analyst to be aware of when they are using the data for research. Include relevant citations.*

The following steps were carried out by the UW capstone team to prepare the data for visualization and synthesis:

1. Created Reading and Filtering Computer Script
   1. Filtering out expired and non-issued permits.
      1. filter(PermitStatus == "Issued")
      2. filter(DateIssued >"2012-01-01")
      3. filter(DateExpired >= Sys.Date())
   2. Filtering for specific and relevant permit types - NMFS 10a1A Salmon, 4d, NMFS BiOp DTA, Tribal 4d.
      1. filter(ResultCode == c("NMFS 10a1A Salmon","4d", "NMFS BiOp DTA", "Tribal 4d"))
   3. Recoded all smolts, fries, larva to “juveniles” and subadults to “adults”.
      1. mutate(LifeStage = recode(LifeStage,

"Smolt" = "Juvenile",

"Fry" = "Juvenile",

"Larvae" = "Juvenile",

"Subadult" = "Adult"))

* 1. Recoded HUC numbers based on reclassification by USGS
     1. mutate(HUCNumber = recode(HUCNumber,

`18020103` = 18020156,

`18020109` = 18020163,

`18020112` = 18020154,

`18020118` = 18020154,

`18040005` = 18040012,

`18060001` = 18060015,

`18060012` = 18060006))

* 1. Recoded and categorized species by their distinct population segments or evolutionary significant unit.
     1. mutate(Species = paste(Population, CommonName, sep = " "))
  2. Recoded and categorized species by production - Natural, Listed Hatchery, Unlisted Hatchery.
     1. mutate(Prod = recode(Production,

"Natural" = "Natural",

"Listed Hatchery" = "Listed Hatchery",

"Listed Hatchery, Clipped and Intact" = "Listed Hatchery",

"Listed Hatchery Intact Adipose" = "Listed Hatchery",

"Listed Hatchery Adipose Clip" = "Listed Hatchery",

"Unlisted Hatchery" = "Unlisted Hatchery"))

* 1. Filtered out non-invasive, non-lethal take actions (i.e. Observe/Harass and Sampling dead animals)
     1. filter(TakeAction != c( "Observe/Harass", "Observe/Sample Tissue Dead Animal, "N/A", "NA"))

1. Developing new fields within data files.
   1. Created a new field to display ESU or DPS under “Species”
      1. mutate(Species = paste(Population, CommonName, sep = " "))
   2. Created a new field to recode and condense production type under “Prod”
      1. mutate(Prod = recode(Production,

"Natural" = "Natural",

"Listed Hatchery" = "Listed Hatchery",

"Listed Hatchery, Clipped and Intact" = "Listed Hatchery",

"Listed Hatchery Intact Adipose" = "Listed Hatchery",

"Listed Hatchery Adipose Clip" = "Listed Hatchery",

"Unlisted Hatchery" = "Unlisted Hatchery"))

* + 1. filter(Prod != "Unlisted Hatchery")
  1. Created a new field to report total mortality that occurred.
     1. create\_totalmorts() - function included in *NMFSResPermits* package
  2. Created a new field to report total authorized take under “AuthTake.
     1. mutate(AuthTake = ExpTake + IndMort)

1. Map Component
   1. Setting up reactive data table (PreAppCode-1)
      1. Used functions from NMFSResPermits package to create new fields and organize data.
      2. Created a location field for the table. This is based on a hierarchical sort of different location based fields from the raw data. When researchers apply for permits they often use different fields to describe their location. The highest level (largest in terms of geographic scope) is used unless it is an NA value, then the next highest non-NA value is used. This hierarchy is WaterbodyName > BasinName > StreamName > LocationDescription.
      3. Created a water type column that uses keywords to identify a given location as freshwater or saltwater.
   2. Setting up data frames for display in map (PreAppCode-2)
      1. Manually assigned HUCs to some marine areas in Washington (Puget Sound and Strait of Juan de Fuca area).
      2. Aggregated both total take and lethal take data, and then joined to spatial data using HUC 8 codes.
   3. Creating ESU/DPS boundaries (PreAppCode-3)
      1. Combine HUC assignment data with spatial data, and then loop through each ESU/DPS and join together all HUC polygons where that ESU/DPS can be found. End result is one large polygon showing the ranges for each ESU/DPS.
2. Time Series Component
   1. Setting up data - Changing time factor
      1. a <- as.factor(wcr\_act$DateReportPeriodEnd)
      2. b<-strptime(a,format="%Y-%m-%d")
      3. Year <- format(as.Date(b, format ="%Y-%m-%d" ), "%Y")
      4. Year <- as.data.frame(Year)
      5. wcr.v <- cbind(wcr\_act, Year) #look into annual time start vs annual start end
   2. Setting up data - Creating totalmorts and QC NA values
      1. wcr.v <-wcr.v %>%

create\_totalmorts() %>%

order\_table() %>%

replace\_na(list(ExpTake = 0, ActTake = 0, TotalMorts = 0, ActMort = 0))

* 1. Stacked bar plot code
     1. Aggregating Authorized Take and Renaming Columns
        1. df <- aggregate(wcr.v$ExpTake, by = list(wcr.v$CommonName, wcr.v$ResultCode, wcr.v$ActMort, wcr.v$ActTake, wcr.v$TakeAction, wcr.v$Species, wcr.v$LifeStage, wcr.v$Prod, wcr.v$Year, wcr.v$TotalMorts), FUN = sum)
        2. names(df) <- c("CommonName", "ResultCode", "ActMort", "ActTake", "TakeAction", "ESU", "LifeStage", "Production", "Year", "TotalMorts", "ExpTake")
     2. Summing each variable by year and grouping by ESU, Production, and Lifestage
        1. YT <-df %>% group\_by(Year, ESU, Production, LifeStage) %>% summarize(Used = sum(ActTake))
        2. YM <- df %>% group\_by(Year, ESU, Production, LifeStage) %>% summarize(Used = sum(ActMort))
        3. TM <- df %>% group\_by(Year, ESU, Production, LifeStage) %>% summarize(Authorized\_Mortality = sum(TotalMorts))
        4. ET <-df %>% group\_by(Year, ESU, Production, LifeStage) %>% summarize(Authorized\_Take = sum(as.numeric(ExpTake)))
     3. Merging data sets
        1. Take <- merge(YT, ET, by = c("Year", "ESU", "Production", "LifeStage"))
        2. Mort <- merge(YM, TM, by = c("Year", "ESU", "Production", "LifeStage"))
     4. Replacing NaN or NA or Inf with 0
        1. Take[is.na(Take)] <- 0
        2. Mort[is.na(Mort)] <- 0
     5. Math for determining unused take
        1. Take %>% mutate(Unused = Authorized\_Take - Used) -> t
        2. Mort %>% mutate(Unused = Authorized\_Mortality - Used) -> m
     6. Delineating columns we want to include within plot data
        1. df1 <- t %>% gather("Take\_Type","N", 5:7)
        2. df2 <- m %>% gather("Take\_Type","N", 5:7)
     7. Creating plot data for further filtering
        1. df1 %>% filter(Take\_Type %in% c("Used","Unused")) -> df\_plot
        2. df2 %>% filter(Take\_Type %in% c("Used","Unused")) -> df\_plot2
  2. Table code
     1. Aggregating Authorized Take and Renaming Columns
        1. dt <- aggregate(wcr.v$ExpTake, by = list(wcr.v$FileNumber, wcr.v$CommonName, wcr.v$ResultCode, wcr.v$ActMort, wcr.v$ActTake, wcr.v$TakeAction, wcr.v$Species, wcr.v$LifeStage, wcr.v$Prod, wcr.v$Year, wcr.v$TotalMorts, wcr.v$CaptureMethod, wcr.v$ReportID), FUN = sum)
        2. names(dt) <- c("FileNumber","CommonName", "ResultCode", "ActMort", "ActTake", "TakeAction", "ESU", "LifeStage", "Production", "Year", "TotalMorts", "CaptureMethod", "ReportID","ExpTake")
     2. Summing each variable by year
        1. YT <-dt %>% group\_by(Year, ESU, Production, LifeStage, FileNumber, CaptureMethod, ReportID, ResultCode) %>% summarize(Reported\_Take = sum(ActTake))
        2. YM <- dt %>% group\_by(Year, ESU, Production, LifeStage,FileNumber, CaptureMethod, ReportID, ResultCode) %>% summarize(Reported\_Mortality = sum(ActMort))
        3. TM <- dt %>% group\_by(Year, ESU, Production, LifeStage, FileNumber, CaptureMethod, ReportID, ResultCode) %>% summarize(Authorized\_Mortality = sum(TotalMorts))
        4. ET <-dt %>% group\_by(Year, ESU, Production, LifeStage, FileNumber, CaptureMethod, ReportID, ResultCode) %>% summarize(Authorized\_Take = sum(as.numeric(ExpTake)))
     3. Merging data sets back together
        1. Take <- merge(YT, ET, by = c("Year", "ESU", "Production", "LifeStage", "FileNumber", "CaptureMethod", "ReportID", "ResultCode"))
        2. Mort <- merge(YM, TM, by = c("Year", "ESU", "Production", "LifeStage", "FileNumber", "CaptureMethod", "ReportID", "ResultCode"))
        3. dt <- merge(Take, Mort, by = c("Year", "ESU", "Production", "LifeStage", "FileNumber", "CaptureMethod", "ReportID", "ResultCode"))
     4. Math for determining the Unused Take/Mortality
        1. dt %>% mutate(Authorized\_Take\_Unused = Authorized\_Take - Reported\_Take) %>% mutate(Authorized\_Mortality\_Unused = Authorized\_Mortality - Reported\_Mortality) -> dt
  3. Extras
     1. Creating custom color palette
        1. mycols <- colors()[c(461, 142, 525, 87)]

4.3. Quality assurance and quality control. *Note anything the data creators did to ensure the completeness and accuracy of their dataset (e.g., instrument calibrations, automated procedures, manual/visual tests for outliers).*

For the purpose of this project, some fields and data entries were modified to simplify analyses and provide consistency across the nomenclature. These include:

* Adjusting HUC 8 codes to encompass redrawn boundaries.
* Renaming and classifying water bodies in the 'WaterbodyName' field to allow for consistent nomenclature. Renaming practices were performed using best available data provided in the 'LocationDescription' field.

4.4. Data completeness and constraints. *Were any data excluded from the dataset? If so, why? What are known cautions or problems, such as sampling bias? Are there ways the data should not be used (according to the data creators)?*

**Assumptions/Limitations**

1. *99999999 and Incorrect HUC Codes* - The input of incorrect HUC 8 codes by researchers applying for permits (for example: the ‘9999999’ HUC codes) meant these data could not be accurately plotted or included, therefore serves as a limitation of this project.
2. *Changed HUC numbers -* Over time, HUC 8 codes have been rearranged and their boundaries redrawn. Many permits use old or outdated HUC 8 codes, causing issues when trying to map our permit data using the Watershed Boundary Dataset (USGS et al. 2022). Therefore, HUC 8 codes had to be updated to reflect any changes to their boundaries. Decisions were made using an unpublished document that summarizes HUC 8 code changes up until 2018 (Hanson et al. 2018). Some of these changes ran on assumptions using other fields (such as WaterbodyName or LocationDescription) and are detailed below:
   1. # 18020103 = 18020156 # very certain
   2. # 18020109 = 18020163 # very certain
   3. # 18020112 = 18020154 # very certain based on location descriptions
   4. # 18020118 = 18020154 # very certain based on location descriptions
   5. # 18040005 = 18040012 # very certain based on location descriptions
   6. # 18060001 = 18060015 # split between 18050006 as well, arbitrarily picked
   7. # 18060012 = 18060006 # chose this over Monterey Bay as population is South-Central Cal Coast
3. *Non-reported take -* Under the data file *WCRPermitBiOp\_Pass report data 4d and S10\_22March22.csv,* researchers are asked to report the actual take and mortality that occurred to ESA-listed species during the duration of their study. However, some organizations or projects can neglect to report on the actual take and mortality. As a result, these projects are not included in our analyses due to lack of data. Thus, a limitation of this unreported data is we are missing total take and mortality data and therefore complicates further abundance data analysis.
4. *Tribal 4d permits* - Tribal 4d permits were omitted from tables showing individual permit information for data privacy reasons, but included in the totals shown in the map and time series. When using this data, be wary of these differences in counts due to this omission.
5. *Omission of “Unknown,” “Unlisted,” or “N/A” values -* in fields where input was necessary, inputs that included “Unknown,” “Unlisted,” or “N/A” were purposefully omitted. This was done to ensure data accuracy and includes fields such as Production and Take Action.

**5. CITATIONS**

1. U.S. Geological Survey (USGS), U.S. Department of Agriculture – Natural Resource Conservation Service (NRCS), U.S. Environmental Protection Agency (EPA) (2022). USGS National Watershed Boundary Dataset in FileGDB 10.1 format (published 20220526). Accessed May 15, 2022 at URL<https://prd-tnm.s3.amazonaws.com/index.html?prefix=StagedProducts/Hydrography/WBD/National/GDB/>
2. Hanson, K., Daw, S., Davenport, L., Jones, K., Niknami, L., & Buto, S. (2018). Criteria for Legacy Name and Code Changes. [Unpublished]