Description of raw CFEC data processing for NCEAS portfolio effects working group

Jordan Watson & Eric Ward

March 16, 2016 (Update 4/7/2016)

Objective

The objective of this document is to summarize the data cleaning that was done to clean the raw CFEC data (1985 - 2014) for use by the NCEAS catch portfolio working group.

Raw data summary

We'll initially restrict our analysis to data after 1984, since the data from 1975-1984 were dominated by foreign vessels (data we don't have). For more information, see AFSC technical reports (such as http://www.afsc.noaa.gov/Publications/AFSC-TM/NOAA-TM-AFSC-54.pdf). The cutoff of 1984/1985 is also consistent with the ADFG use. Further, the species reporting changes drastically before / after 1985 – examples include no rockfish (demersal, pelagic, slope categories below), flatfish (shallow, deep categories below).

Additional information on CFEC data here,
Gear descriptions: https://www.cfec.state.ak.us/misc/GEARDESC.HTM
Fishery codes: https://www.cfec.state.ak.us/misc/FshyDesC.htm
Species and gear codes: https://www.adfg.alaska.gov/static/license/fishing/pdfs/elanding_codes.pdf
Commercial areas: http://www.adfg.alaska.gov/index.cfm?adfg=fishingCommercialByFishery.statmaps
Halibut areas: http://www.iphc.int/publications/techrep/tech0049.pdf

The file "cfec_w_IDs" is a zip file of an R workspace created by Jordan Watson on 03/15/2016. This file is a large workspace (~ 15 GB) that contains a single names object ('cfecnew'). The size of the data frame is approximately 19.8 million rows by 77 variables.

```
setwd("//nmfs.local/AKC-ABL/Users2/jordan.watson/Desktop/AFSC/GOA/AKFIN/Meeting")
load("Data/cfec_w_IDs.RData")
cfec = cfecnew
rm(cfecnew) # remove the old named object

#idx = grep("S 04T", cfec$p_fshy)
#group_by(cfec[idx,], year) %>%
#  summarize(nSetnet = length(unique(p_holder)))
# Remove A80 and CP vessels
A80<-c(68870,29923,22011,39798,59870,8500,41219,62545,56965,54693,1119,57228,69038,61081,61083,55921,57
CP <-c(56987,59378,57621,59170,55767,57450,54886,56991,60407,60795,60202,60660,56789,56618,59503)
cfec <- filter(cfec,!adfg %in% c(A80,CP))
rm(A80,CP)
```

For clarity, this workspace contains some variables that have been renamed. The original variable names were changed with the following commented out code,

```
#cfecnew <- rename(cfecnew, year=F_YEAR,          # ADFG_B_BATCH_YEAR
#                  port=F_PORT,          # ADFG_H_PORT_CODE
```

```
#                p_holder=P_FILENO,    # CFEC_FILE_NUMBER - unique permit holder ID
#                p_serial=P_SERIAL,    # CFEC_PMT_SERIAL  - unique permit ID
#                g_earn=G_EARN,        # CFEC_VALUE
#                g_pounds=F_NETLBS,    # ADFG_I_POUNDS
#                f_ticket=F_TICKET,    # ADFG_H_SEQ_TICKET_NUMBER - unique fish tiket ID
#                adfg=F_ADFG,          # ADFG_H_ADFG_NUMBER - don't use this one
#                cadfg=C_ADFG,         # CFEC_CORRECTED_ADFG
#                harvest=F_HARVST,     # ADFG_I_HARVEST_CODE
#                procid=F_PROCSS,      # ADFG_H_PROCESSOR_CODE
#                g_spcs=G_SPCS,        # CFEC_SPECIES_CODE
#                stat6=F_STAT,         # CFEC_STAT_AREA
#                specn=F_SPCS,         # ADFG_I_SPECIES_CODE
#                p_fshy=P_FSHY,        # CFEC_PMT_FSHY
#                p_check=P_CHECK)      # CFEC_PMT_CHECK

# cfecnew$g_price=ifelse(cfecnew$g_pounds>0,cfecnew$g_earn/cfecnew$g_pounds,0)
```

Deleting variables

There's a number of variables that can be deleted either because they have no information (null values), have been used inconsistently (the double permit variables 'D_OPER', etc. were used 2008-2013), or are either redundant or not informative with respect to the questions we're interested in, so we'll drop these out to save memory. More details about these variables can be found in the associated metadata (.xlsx workbook).

```
isNull = c("F_FG_MGT","C_CFACT", "F_DISTRC", "F_ERRFLG", "F_AYK", "G_FED_AR","F_BTAREA",
"A_CITY", "A_STATE", "A_ZIP", "P_YEAR", "I_NAME", "I_NAMTYP", "F_POT_DR", "F_PRCENT",
"C_ADFGST", "F_CTDATE", "F_LNDATE", "F_NUMBER", "F_VALUE", "F_POUNDS", "F_TNUMB",
"F_BATCH", "F_PERIOD",  "F_EMBSEQ", "F_PROCCK", "F_WEEK",  "F_TTYPE",
"F_ITEM", "F_FSHERY",  "G_GEAR", "G_AREA", "G_PRC_AR",  "P_YEAR", "P_PMTSEQ",
"p_check", "P_ID", "P_FEETYP", "P_ADFG", "P_FEEFLG", "P_FSHBLE",
"F_MPGMNM", "D_OPER", "D_O_FSHY", "D_O_PNUM", "D_O_PCHK", "D_O_YSEQ", "F_SIZE",
"D_O_YCHK", "adfg")
cfec = cfec[,-which(names(cfec)%in%isNull)]
rm(isNull)
```

This significantly reduces the size of the dataset, to the remaining 29 variables,

```
names(cfec)
```

```
##  [1] "F_REGION" "year"     "g_pounds" "F_PERMIT" "F_INPFC"  "F_GEAR"
##  [7] "port"     "procid"   "f_ticket" "F_DELIVR" "harvest"  "specn"
## [13] "stat6"    "g_spcs"   "g_earn"   "p_serial" "p_holder" "p_fshy"
## [19] "P_TYPE"   "P_STATUS" "A_RES"    "cadfg"    "C_LANDST" "F_DISPOS"
## [25] "F_MPGMID" "startdt"  "landdate" "g_price"
```

The fish ticket field (f_ticket) provides unique fish ticket numbers by year. To make unique fish tickets across all years, affix the last two digits of each year to the fish ticket ID number. We subesquently overwrite the original column with the new one.

```
cfec <- group_by(cfec,year) %>% mutate(ft=paste0(gsub("19","",year),f_ticket,p_holder)) %>% dplyr::sele
```

Deleting records

The CFEC data contains a number of records that we don't want to be included in our analyses. The first obvious group includes landings that are non-commercial, from test-fisheries, discards, etc. There are several fields we can use to do this.

First, we'll use 'C_LANDST' to remove all non-commercial records. We won't want to use commercial landings that are bycatch (Pacific Salmon Commission) or discards, because they'd show up as having no value. As a second filter, we can use the corrected ADFG number 'cadfg' to remove those problematic records (missing values, values < 10000, values > 98000). Third, we'll delete when the 'P_STATUS' is blank or 'O' (15 records). We may want to filter on the remaining codes,
A: revoked permit holder
C: current permit holder
E: former emergency transferor
F: former permit holder (permanent transferor)
I: inactive emergency transferee

Finally, we can use the 'P_TYPE' field to include the following 4 types of permits:
E: interim-entry in a limited fishery
I: interim-use in an unlimited fishery
M: moratorium
P: permanent in a limited fishery

```
cfec = cfec[which(cfec$C_LANDST == "C"), ]

#cfec = cfec[-which(cfec$cadfg == ""), ]
#cfec = cfec[-which(cfec$cadfg < 10000 | cfec$cadfg >= 98000), ]

cfec = cfec[-which(cfec$P_STATUS%in%c("", "O")), ]

#table(cfec$P_STATUS)
```

Merging species and gears

Gears

We have to clean up a few of the gears. Drop a few gears that are sporadically used, and combine longline and pot gears because of changes in how those gears were reported. This doesn't destroy information, because we still have data on vessel length, etc.

```
cfec$gearn = as.numeric(cfec$F_GEAR)
cfec = cfec[, -which(names(cfec)=="F_GEAR")]
# 8 = fishwheel, used at start / end, 676 recs
# 13 = dipnet, only last 2 years, 329 recs
# 23 = Hydro/mechanical dredge, only 74 recs
# 27 = Double otter trawl, 37 recs
# 37 = Pair trawl, 16 recs
# 41 = Sunken gillnet, 59 recs
# 85 = ??, 5 recs
# 90 = trap, 1 rec
# 99 = other, stopped reporting, 175 recs
cfec = cfec[-which(cfec$gearn%in%c(8, 13, 23, 27, 37, 41, 85, 90, 99)),]
# Merge 6 and 61 into longline
cfec$gearn[which(cfec$gearn==61)] = 6
```

```
# Merge all pot gears together -- note only 9 and 91 are reported
cfec$gearn[which(cfec$gearn%in%c(9, 19, 29, 39, 49, 59, 69, 91))] = 9
```

Species

Because of changes in species reporting (examples = rockfish, the use of unidentified species), we need to do some lumping some species. Groupings were consistent to be consistent with the ADFG codes, so for example rockfish are grouped into a 'pelagic', 'demersal', and 'slope' group.

```
cfec$specn.orig = cfec$specn

# Lump all skates together
cfec$specn[cfec$specn==700] = 700
cfec$specn[cfec$specn==701] = 700
cfec$specn[cfec$specn==702] = 700
cfec$specn[cfec$specn==703] = 700
cfec$specn[cfec$specn==704] = 700
cfec$specn[cfec$specn==705] = 700

# Lump all sharks together
cfec$specn[cfec$specn==689] = 689
cfec$specn[cfec$specn==690] = 689
cfec$specn[cfec$specn==691] = 689
cfec$specn[cfec$specn==692] = 689

# Lump flatfish (Shallow)
cfec$specn[cfec$specn==123] = 119 # rock
cfec$specn[cfec$specn==129] = 119 # starry flounder
cfec$specn[cfec$specn==127] = 119 # yellowfin
cfec$specn[cfec$specn==126] = 119 # butter
cfec$specn[cfec$specn==122] = 119 # flathead
cfec$specn[cfec$specn==125] = 119 # rex
cfec$specn[cfec$specn==128] = 119 # English
cfec$specn[cfec$specn==131] = 119 # petrale
cfec$specn[cfec$specn==132] = 119 # sand
cfec$specn[cfec$specn==133] = 119 # Alaska plaice
cfec$specn[cfec$specn==116] = 119 # Bering flounder
cfec$specn[cfec$specn==117] = 119 # Kamchatka flounder

# 5, 6, 7, 15, 26, 47, 61,
# Lump flatfish (deep)
cfec$specn[cfec$specn==124] = 118 # dover
cfec$specn[cfec$specn==134] = 118 # greenland turbot

# Lump rockfish, codes from here:
# Slope rockfish
cfec$specn[cfec$specn==136] = 144 # northern
cfec$specn[cfec$specn==137] = 144 # bocaccio
cfec$specn[cfec$specn==141] = 144 # POP
cfec$specn[cfec$specn==143] = 144 # thornyhead
cfec$specn[cfec$specn==144] = 144 # unid
cfec$specn[cfec$specn==151] = 144 # rougheye
cfec$specn[cfec$specn==152] = 144 # shortraker
```

```
cfec$specn[cfec$specn==153] = 144 # redbanded
cfec$specn[cfec$specn==157] = 144 # silvergray
cfec$specn[cfec$specn==158] = 144 # redstripe
cfec$specn[cfec$specn==159] = 144 # darkblotched
cfec$specn[cfec$specn==166] = 144 # sharpchin
cfec$specn[cfec$specn==176] = 144 # harlequin
cfec$specn[cfec$specn==177] = 144 # blackgill
cfec$specn[cfec$specn==179] = 144 # pygmy
cfec$specn[cfec$specn==181] = 144 # shortbelly
cfec$specn[cfec$specn==182] = 144 # splitnose
cfec$specn[cfec$specn==183] = 144 # stripe tail
cfec$specn[cfec$specn==184] = 144 # vermillion
cfec$specn[cfec$specn==185] = 144 # aurora

# Demersal rockfish
cfec$specn[cfec$specn==135] = 168 # greenstripe
cfec$specn[cfec$specn==138] = 168 # copper
cfec$specn[cfec$specn==139] = 168 # unid
cfec$specn[cfec$specn==140] = 168 # red
cfec$specn[cfec$specn==145] = 168 # yelloweye
cfec$specn[cfec$specn==146] = 168 # Canary
cfec$specn[cfec$specn==147] = 168 # quillback
cfec$specn[cfec$specn==148] = 168 # tiger
cfec$specn[cfec$specn==149] = 168 # China
cfec$specn[cfec$specn==150] = 168 # rosethorn
cfec$specn[cfec$specn==171] = 168 # shortraker/rougheye
cfec$specn[cfec$specn==175] = 168 # yellowmouth
cfec$specn[cfec$specn==178] = 168 # chilipepper

# Pelagic rockfish
cfec$specn[cfec$specn==142] = 169 # black
cfec$specn[cfec$specn==154] = 169 # dusky
cfec$specn[cfec$specn==155] = 169 # yellowtail
cfec$specn[cfec$specn==156] = 169 # widow
cfec$specn[cfec$specn==167] = 169 # blue
cfec$specn[cfec$specn==172] = 169 # dusky
cfec$specn[cfec$specn==173] = 169 # dark

# Group sculpins
cfec$specn[cfec$specn %in% c(160:165)] = 160

# Group greenlings (not AMCK or lingcod)
cfec$specn[cfec$specn %in% c(190,191,192,194)] = 190

# Group all osmerids together
cfec$specn[cfec$specn %in% c(510:516)] <- 510

# Group all char
cfec$specn[cfec$specn %in% c(520:532)] <- 520  # char

# Group all trout
cfec$specn[cfec$specn%in%c(540:562)] <- 540  # trout
```

```
#  Whitefish and trout did not have text labels in the species database.
#  We created our own for easier reference but given this lack of detail
#  it seems reasonable to lump these species groups for the sake of clarity
# Whitefish
cfec$specn[cfec$specn==580] <- 580  # Whitefish
cfec$specn[cfec$specn==583] <- 580  # Least cisco
cfec$specn[cfec$specn==584] <- 580  # Arctic cisco
cfec$specn[cfec$specn==585] <- 580  # Bering cisco
cfec$specn[cfec$specn==589] <- 580  # Humpback whitefish

# Halibut - appear under several different historic species codes so we consolidate these
cfec$specn[cfec$specn %in% c(200:204)] <- 200

# Chinook have two species codes. Consolidate.
cfec$specn[cfec$specn==411] <- 410

# Salmon roe (by species) get lumped into a single salmon roe category. This was a topic of much discus
cfec$specn[cfec$specn %in% c(401:405)] <- 400
cfec$specn[cfec$specn%in%c(410:450) & cfec$F_DELIVR==14] <- 400

# Herring are listed under 5 different species codes but 231:235 only account for a few thousand bucks.
# We'll lump them into the sac-roe species code because they will have no effect on the numbers. But th
# of making a bunch of extra ones.
cfec$specn[cfec$specn %in% c(231:235)] <- 23043

# Most herring landings occur under species 230, with different F_DELIVR codes
# Create a species code for the sac-roe fishery
cfec$specn[cfec$specn==230 & cfec$F_DELIVR==43] <- 23043
# Create a species code for the bait/food fishery
cfec$specn[cfec$specn==230 & cfec$F_DELIVR==44] <- 23044
# Create a species code for the pound fishery
cfec$specn[cfec$specn==230 & cfec$F_DELIVR==45] <- 23045
# There may be 1 or 2 records with other codes, and which probably have no dollar value.
# Lump them into the sac roe fishery to avoid having extra codes
cfec$specn[cfec$specn==230] <- 23043

# Lump all king crabs
cfec$specn[cfec$specn %in% c(920:925)] <- 920

# Lump all king crabs
cfec$specn[cfec$specn %in% c(930:934)] <- 930

# Lump all remaining crabs (not Dungeness)
cfec$specn[cfec$specn %in% c(900,940:953)] <- 940

# Lump all shrimp
cfec$specn[cfec$specn %in% c(960:965)] <- 960

# Lump non-geoduck and non-razor clams together
cfec$specn[cfec$specn %in% c(810,812,820,825,840,842)] <- 810

# Lump weathervane and pink scallops
cfec$specn[cfec$specn %in% c(850:851)] <- 850
```

```
# Lump urchins
cfec$specn[cfec$specn %in% c(892:893,896)] <- 892

# Lump sea cucumbers
cfec$specn[cfec$specn %in% c(894:895)] <- 895

# Create an "OTHER" category that includes misc freshwater and a bunch of random trawl caught spec
cfec$specn[cfec$specn %in% c(100,101,112,170,180,190,206:220,250,260,310:380,490,500,520,540,570,580,590

# We no longer need the F_DELIVR column and can remove it
cfec <- dplyr::select(cfec,-F_DELIVR)

#-----------------------------------------
# To add text species codes, load and join with Species lookup table
load("Data/SpeciesTable.RData")

#  000,033 and 095 don't show up in the species codes (they account for only 7 records)
#  Remove them
cfec <- filter(cfec,!specn %in% c("000","033","095"))
#  To avoid issues with different factor levels, convert specn to numeric for now.
cfec$specn <- as.numeric(as.character(cfec$specn))

#  Join the cfec dataset with the "spec"
cfec <- left_join(cfec,spec.tbl)
```

```
## Joining by: "specn"
```

```
#  Rename some of the groups
cfec$spec <- as.character(cfec$spec)

#  Rename some of the groups
cfec$spec <- as.character(cfec$spec)
cfec$spec[cfec$specn==118] = "FLAT.deep"
cfec$spec[cfec$specn==119] = "FLAT.shallow"
cfec$spec[cfec$specn==144] = "ROCK.slope"
cfec$spec[cfec$specn==168] = "ROCK.demersal"
cfec$spec[cfec$specn==169] = "ROCK.pelagic"
cfec$spec[cfec$specn==206] = "OTHER"
cfec$spec[cfec$specn==400] = "aUROE"
cfec$spec[cfec$specn==689] = "SHARK"
cfec$spec[cfec$specn==700] = "SKATE"
cfec$spec[cfec$specn==810] = "OCLM"
cfec$spec[cfec$specn==850] = "SCAL"
cfec$spec[cfec$specn==892] = "URCH"
cfec$spec[cfec$specn==920] = "KCRB"
cfec$spec[cfec$specn==930] = "TCRB"
cfec$spec[cfec$specn==940] = "OCRB"
cfec$spec[cfec$specn==960] = "USRM"
cfec$spec[cfec$specn==23043] = "SAC"
cfec$spec[cfec$specn==23044] = "BAIT"
cfec$spec[cfec$specn==23045] = "POUND"

cfec$spec[cfec$spec==""] <- NA
```

```
# drop records with no spp
cfec = cfec[-which(is.na(cfec$spec)),]
rm(spec.tbl)
```

Assigning coarse spatial regions

Examining the spatial diversity of fishing behavior is complicated because we have 3 types of area codes. The 6-digit codes represent groundfish management areas (federal), the 5-digit codes are state salmon/shellfish areas, and the 3-digit codes are halibut areas. For the groundfish and salmon areas, there are sub-areas that are also reported (as the last 2-digits), so we'll lump sub-areas together by stripping off the last 2-digits. The challenge too is that all three of these areas overlap in space – so in a single location (latitude, longitude), a vessel may land a fish and record it in any of the 3 areas (for examples of this occuring, see species like sablefish).

For larger spatial comparisons, we'll create a field called 'area'. We'll group groundfish / salmon-shellfish / halibut areas into the following regions: Prince William Sound, Kodiak (east and west), Cook Inlet, and Southeast Alaska. More regions could be added below, and records that don't fit into one of these regions are left as NAs. The justification of these boundaries are largely based on halibut areas (which are the largest spatially) and by a series of k-means cluster analyses that identified similar groupings.

```
cfec$area = NA
cfec$stat6 = as.numeric(cfec$stat6)

cfec$area[cfec$stat6==485900] = "PWS"
cfec$area[cfec$stat6==475900] = "PWS"
cfec$area[cfec$stat6==465900] = "PWS"
cfec$area[cfec$stat6==455900] = "PWS"
cfec$area[cfec$stat6==445900] = "PWS"
cfec$area[cfec$stat6==495930] = "PWS"
cfec$area[cfec$stat6==485930] = "PWS"
cfec$area[cfec$stat6==475930] = "PWS"
cfec$area[cfec$stat6==465930] = "PWS"
cfec$area[cfec$stat6==455930] = "PWS"
cfec$area[cfec$stat6==445930] = "PWS"
cfec$area[cfec$stat6==486000] = "PWS"
cfec$area[cfec$stat6==476000] = "PWS"
cfec$area[cfec$stat6==466000] = "PWS"
cfec$area[cfec$stat6==456000] = "PWS"
cfec$area[cfec$stat6==446000] = "PWS"
cfec$area[cfec$stat6==486030] = "PWS"
cfec$area[cfec$stat6==476030] = "PWS"
cfec$area[cfec$stat6==466030] = "PWS"
cfec$area[cfec$stat6==456030] = "PWS"
cfec$area[cfec$stat6==486100] = "PWS"
cfec$area[cfec$stat6==476100] = "PWS"
cfec$area[cfec$stat6==466100] = "PWS"
cfec$area[which(cfec$stat6 < 20099 & cfec$stat6 >= 20000)] = "PWS"
cfec$area[which(cfec$stat6 < 21299 & cfec$stat6 >= 21200)] = "PWS"
cfec$area[which(cfec$stat6 < 22799 & cfec$stat6 >= 22700)] = "PWS"
cfec$area[which(cfec$stat6 < 22699 & cfec$stat6 >= 22600)] = "PWS"
cfec$area[which(cfec$stat6 < 22899 & cfec$stat6 >= 22800)] = "PWS"
cfec$area[which(cfec$stat6 < 22199 & cfec$stat6 >= 22100)] = "PWS"
```

```
cfec$area[which(cfec$stat6 < 22299 & cfec$stat6 >= 22200)] = "PWS"
cfec$area[which(cfec$stat6 < 22599 & cfec$stat6 >= 22500)] = "PWS"
cfec$area[which(cfec$stat6 < 22499 & cfec$stat6 >= 22400)] = "PWS"
cfec$area[which(cfec$stat6 < 22399 & cfec$stat6 >= 22300)] = "PWS"
# Add halibut areas
cfec$area[cfec$stat6 %in% c(242,232,220,230,240)] = "PWS"

cfec$area[cfec$stat6==525900] = "CookInlet"
cfec$area[cfec$stat6==535900] = "CookInlet"
cfec$area[cfec$stat6==515930] = "CookInlet"
cfec$area[cfec$stat6==525930] = "CookInlet"
cfec$area[cfec$stat6==535930] = "CookInlet"
cfec$area[cfec$stat6==516000] = "CookInlet"
cfec$area[cfec$stat6==526000] = "CookInlet"
cfec$area[cfec$stat6==496030] = "CookInlet"
cfec$area[cfec$stat6==506030] = "CookInlet"
cfec$area[cfec$stat6==516030] = "CookInlet"
cfec$area[cfec$stat6==526030] = "CookInlet"
cfec$area[cfec$stat6==506100] = "CookInlet"
cfec$area[cfec$stat6==516100] = "CookInlet"
cfec$area[which(cfec$stat6 < 24799 & cfec$stat6 >= 24400)] = "CookInlet"
cfec$area[which(cfec$stat6%in%c(261, 272))] = "CookInlet"

cfec$area[which(cfec$stat6 < 11599 & cfec$stat6 >= 11100)] = "Southeast"
cfec$area[which(cfec$stat6 < 10499 & cfec$stat6 >= 10100)] = "Southeast"
cfec$area[which(cfec$stat6 < 11099 & cfec$stat6 >= 10500)] = "Southeast"
cfec$area[which(cfec$stat6 >= 305430 & cfec$stat6 <= 305530)] = "Southeast"
cfec$area[which(cfec$stat6 >= 315430 & cfec$stat6 <= 315600)] = "Southeast"
cfec$area[which(cfec$stat6 >= 325430 & cfec$stat6 <= 325700)] = "Southeast"
cfec$area[which(cfec$stat6 >= 335430 & cfec$stat6 <= 335800)] = "Southeast"
cfec$area[which(cfec$stat6 >= 345400 & cfec$stat6 <= 345830)] = "Southeast"
cfec$area[which(cfec$stat6 >= 355400 & cfec$stat6 <= 355900)] = "Southeast"
cfec$area[which(cfec$stat6 > 365400 & cfec$stat6 < 365830)] = "Southeast"
# For IPHC see Fig. 20 here: http://www.iphc.int/publications/techrep/tech0049.pdf
cfec$area[which(cfec$stat6%in%c(140, 141, 142, 143, 144, 150, 151, 152, 153, 160, 161,
162, 163, 170, 171, 172, 174, 174, 181, 182, 183, 184, 185))] = "Southeast"

# Kodiak east area is defined by halibut areas 270/280, includes ADFG salmon /
# shellfish on E side of kodiak, and groundfish areas extending to NMFS Area 630
# boundary
cfec$area[cfec$stat6%in%c(270,280)] = "Kodiak.east"
cfec$area[which(cfec$stat6 < 25899 & cfec$stat6 >= 25800)] = "Kodiak.east"
cfec$area[which(cfec$stat6 < 25999 & cfec$stat6 >= 25900)] = "Kodiak.east"
cfec$area[cfec$stat6==535630] = "Kodiak.east"
cfec$area[cfec$stat6==525630] = "Kodiak.east"
cfec$area[cfec$stat6==515630] = "Kodiak.east"
cfec$area[cfec$stat6==505630] = "Kodiak.east"
cfec$area[cfec$stat6==495630] = "Kodiak.east"
cfec$area[cfec$stat6==535700] = "Kodiak.east"
cfec$area[cfec$stat6==525700] = "Kodiak.east"
cfec$area[cfec$stat6==515700] = "Kodiak.east"
cfec$area[cfec$stat6==505700] = "Kodiak.east"
cfec$area[cfec$stat6==495700] = "Kodiak.east"
```

```
cfec$area[cfec$stat6==525730] = "Kodiak.east"
cfec$area[cfec$stat6==515730] = "Kodiak.east"
cfec$area[cfec$stat6==505730] = "Kodiak.east"
cfec$area[cfec$stat6==495730] = "Kodiak.east"
cfec$area[cfec$stat6==525800] = "Kodiak.east"
cfec$area[cfec$stat6==515800] = "Kodiak.east"
cfec$area[cfec$stat6==505800] = "Kodiak.east"
cfec$area[cfec$stat6==495800] = "Kodiak.east"

cfec$area[cfec$stat6%in%c(271,281)] = "Kodiak.west"
cfec$area[cfec$stat6==535730] = "Kodiak.west"
cfec$area[cfec$stat6==545730] = "Kodiak.west"
cfec$area[cfec$stat6==525800] = "Kodiak.west"
cfec$area[cfec$stat6==535800] = "Kodiak.west"
cfec$area[cfec$stat6==545800] = "Kodiak.west"
cfec$area[cfec$stat6==535830] = "Kodiak.west"
cfec$area[cfec$stat6==525830] = "Kodiak.west"
cfec$area[which(cfec$stat6 < 25699 & cfec$stat6 >= 25500)] = "Kodiak.west"
cfec$area[which(cfec$stat6 < 25499 & cfec$stat6 >= 25300)] = "Kodiak.west"
cfec$area[which(cfec$stat6 < 26299 & cfec$stat6 >= 26000)] = "Kodiak.west"

# Bristol Bay, defined as going west to the boundary between NMFS areas 512 / 516
cfec$area[cfec$stat6%in%c(31600, 31700, 31800, 32000, 32100,
  32200, 32400, 32500, 32600)] = "BristolBay"
cfec$area[cfec$stat6%in%c(575730, 575800, 575830)] = "BristolBay"
cfec$area[cfec$stat6%in%c(585700, 585730, 585800, 585830)] = "BristolBay"
cfec$area[cfec$stat6%in%c(595630, 595700, 595730, 595800, 595830)] = "BristolBay"
cfec$area[cfec$stat6%in%c(605600, 605630, 605700, 605730, 605800, 605830)] = "BristolBay"
cfec$area[cfec$stat6%in%c(615600, 615630, 615700, 615730, 615800, 615830)] = "BristolBay"

cfec <- as.data.frame(cfec)
```

Add vessel lengths to data

```
#  Load the vessel characteristics data that were created using "Vessel_Characteristics.Rmd."
load("Data/vessel.RData")

#  For now, we only really need the vessel length data but different/additional fields can be added lat

#  Convert fields to make them compatible with the vessel data.
cfec$cadfg2 <- as.numeric(cfec$cadfg)
cfec$year <- as.numeric(cfec$year)

#  I use a right join here instead of left join for continuity since "vessel" was the already loaded da
cfec <- dplyr::select(vessel,year,adfg,VLength) %>% right_join(cfec,by=c("adfg"="cadfg2","year"="year"))

#  Create an index value so that we can link back to the original (unscrambled data) if necessary.
cfec$INDEX <- 1:nrow(cfec)
```

Strip the permit holder's identity

Just for safety (because 'p_holder' is technically traceable to a person), we're going to convert 'p_holder', 'cadfg', 'p_serial' to a non-traceable identifier. We'll also re-number fish tickets 1:n, so there's no information

here that could be used to identify individuals and/or vessels.

```
cfec$p_holder = as.numeric(as.factor(cfec$p_holder))
cfec$cadfg = as.numeric(as.factor(cfec$cadfg))
cfec$p_serial = as.numeric(as.factor(cfec$p_serial))

# number fish tickets sequentially
cfec$f_ticket = as.numeric(as.factor(cfec$f_ticket))
cfec$F_PERMIT = as.numeric(as.factor(cfec$F_PERMIT))
```

Output data file for use on other projects

```
#save(cfec, file = "cfecCleaned2.Rdata")
```

Meta-data for saved variables

F_REGION

year Year of landed catch

g_pounds This was *F_NETLBS* in original CFEC data. Net pounds what gets delivered to dock.

F_PERMIT permit fishery, serial, then check digit.

F_INPFC An identification number assigned to a CDQ group by NMFS. Only applicable if we're interested in CDQ quota.

port This was *F_PORT* in original CFEC data. Port code.

procid This was *F_PROCSS* in original CFEC data. Processor code.

f_ticket This fish ticket identification number is stamped on the fish tickets at the ADFG local area offices and is annually unique. First 2 digits = office #, last 6 = serial. Our data only have 6 digits total and I'm not sure why. However, in order to make values that are unique across all years and not just within a year, we have appended the last two digits of the year to the front of the ticket.

harvest This was *F_HARVST* in original CFEC data. This code is used to separate the common property commercial catch from a variety of specialty fisheries. Examples of specialty fisheries are test fisheries, educational fisheries, hatchery controlled fisheries, and derbies.

specn This was *F_SPCS* in original CFEC data. Species number. As a note of caution, care should be taken in merging our cleaned databases with other datasets because of how we changed the species codes in some cases (e.g. all halibut 200-204 became 200, or see rockfish groups above).

specn.orig The original version of the specn field before species were lumped into categories.

spec A text code created for each species. Some of these are the original AKFIN names and others were created by us. See the Species_Categories.xls file for descriptions.

stat6 This was *F_STAT* in original CFEC data. The 3, 5, or 6 digit ADF&G statistical code denoting a specific area of catch.

g_spcs G_SPCS in original CFEC data. A species code assignment based on grouped ADF&G species codes (*ADFG_I_SPECIES_CODE*). 'A' is abalone, 'B' is halibut, 'C' is sablefish, 'D' is Dungeness crab, 'E' is hair crab, 'F' is fresh water fish, 'G' is sac roe herring, 'H' is herring, bait herring, and food herring, 'I' is lingcod, 'J' is geoduck, 'K' is king crab, 'L' is herring roe or herring roe on kelp or other substrate, 'M'

is groundfish, 'N' is snail, 'O' is octopus or squid, 'P' is shrimp, 'Q' is sea cucumber, 'R' is clam (except geoduck), 'S' is salmon, 'T' is Tanner cab, 'U' is sea urchin, 'W' is scallops, 'Y' is Southeast Alaska demersal shelf rockfish, and 'Z' are miscellaneous marine invertebrates.
g_earn G_EARN in original CFEC data. Estimated ex-vessel value (gross earnings) of a fish ticket item based on the net pounds (ADFG_I_POUNDS) multiplied by the price per pound.

p_serial P_SERIAL in original CFEC data. Permit serial number. Different serial number ranges are used depending upon the permit type (*CFEC_PMT_TYPE*). Mariculture (aquatic farm) is 10000-10499, experimental is 10500-10599, test fishing is 10600-10699, educational is 10700-10799, reservation is 10800-10899, hatchery cost recovery is 10900-10999, interim-use in an unlimited fishery is 11000-49999, interim-entry in a limited fishery is 50000-54999, moratorium is 50000-54999, vessel moratorium is 50000-54999, permanent in a limited fishery is 55000-99999, vessel permanent in a vessel limited fishery is 55000-99999.

p_holder This was *P_FILENO* in original CFEC data. A unique number assigned to each permit holder. These file numbers are generated from the ID number and range from 000000 to 999999.

p_fshy P_FSHY in original CFEC data. Permit fishery code. The 6 byte fishery code consists of a 2 byte resource code; a 2 byte gear code or a 1 byte gear code and a 1 byte vessel restriction or 1 byte gear code and 1 byte gear restriction; and a 1 byte CFEC administrative area. The 6th byte indicates a landing permit ('L'), vessel permit ('V'), or CDQ group ('A', 'B', 'C', 'D', 'E', 'F', or 'G').

P_TYPE The type of permit. 'F' is a mariculture (aquatic farm) permit, 'X' is experimental, 'T' is test fishing, 'C' is educational, 'R' is reservation, 'H' is hatchery cost recovery, 'I' is interim-use in an unlimited fishery, 'E' is interim-entry in a limited fishery, 'M' is moratorium, 'V' is vessel moratorium, 'P' is permanent in a limited fishery, and 'L' is permanent vessel permit in a vessel limited fishery.

P_STATUS Permit holder status. 'A' is revoked permit holder, 'C' is current permit holder, 'D' is current permit holder that emergency transferred permit (transferor, valid in current year only), 'E' is former emergency transferor, 'F' is former permit holder (permanent transferor), 'I' is inactive emergency transferee, 'O' is old permit owner (permit never active, this indicates who paid renewal fees when a transfer is done after fees have been paid but before the calendar year of the permit), 'T' is current permit holder of emergency transfer (transferee, valid in current year only).

A_RES The declared residency of the permit holder. 'R' is Alaska resident, 'N' is nonresident, and 'U' indicates the person has never signed a statement regarding residency.

cadfg This was *C_ADFG* in original data. Corrected ADFG code.

C_LANDST A flag indicating whether or not the landing is defined as commercial or non-commercial catch by CFEC. 'C' indicates commercial catch and 'N' indicates non-commercial catch.

F_MPGMID Indicates the program under which fishing occurred. It is the management program abbreviation such as LE (limited entry), CDQ, AFA, or ADAK. For IFQ crab, halibut, or sablefish, the management program is IFQ. For groundfish with no other management program the value should be OA, for open access. Populated 2005 on in shellfish and groundfish. Populated 2009 on for Salmon database data with landing report number.

startdt The original data had a field *F_CTDATE* - Date gear was deployed in the water or harvest of fish or shellfish was initiated. This field was only in the MMDD format. startdt is a concatenation of the *F_CTDATE* and the year of fishing *F_YEAR*.

landdate The original data had a field *F_LNDATE* - The date fish or shellfish are off-loaded or trans-shipped from the catcher vessel to the first purchaser. The land date is opposed to Catch Date however; these two dates can be the same. If off-loading takes more than one day to complete, the land date is the day off-load is completed. For catcher-seller vessels the land date is the date the product is initially. This field was only in the MMDD format, so landdate is a concatenation of the *F_LNDATE* and the year of fishing *F_YEAR*.

g_price Calculated as g_earn/g_pounds

gearn Numeric code used to represent gears.

spec Derived species code. This should be used for coarse filtering, but metrics of diversity should be calculated from the numeric version, *specn*.

area Coarse areas for comparison across the Gulf of Alaska.