Permits, permit holders, vessels 2

Jordan Watson

March 9, 2016

This file summarizes the cfec catch data into a few different objects. These objects contain p_holder annual summaries of landed pounds and earnings by gear type, area fished and by species. Both wide and long formats are provided for easier modeling / plotting. The output of this file is an RData image called "cfecModelsetup.RData."

```
load("Data/cfecCleaned_new.RData")

# initially only use data 1985+
cfec = cfec[cfec$year > 1984, ]
cfec <- rename(cfec, adfg = cadfg) %>% data.frame
cfec$year <- as.numeric(cfec$year)

# Read-in the CPI adjusted for 2009 base dollars.
cpi <- read.csv("Data/CPI_2009.csv") %>% filter(year > 1984)
cfec <- left_join(cfec, cpi)
# Adjust g_earn and g_price to 2009 dollars
cfec[, c("g_earn", "g_price")] <- cfec[, c("g_earn", "g_price")]/cfec$GDPDEF09_base1
cfec$GDPDEF09_base1 <- NULL
rm(cpi)
```

```
# Get rid of the spaces in p_fshy
cfec$p_fshy <- gsub(" ", "", cfec$p_fshy)

# Read in the fishing locations by fishery permit (p_fshy)
p_fshy <- read.csv("Data/new.locs.csv")
p_fshy$p_fshy <- as.character(p_fshy$p_fshy)

# Combine these areas (original and our coarse new area) to the data
cfec <- left_join(cfec, p_fshy)

# Now for gears...  Read in gear description file and link with cfec.
gears <- read.csv("Data/new_gear_cat.csv")
cfec <- left_join(cfec, gears[, c("gearn", "Gear")])

# Troll fisheries only happen in Southeast
cfec$New.Loc[cfec$New.Loc == "Statewide" & cfec$Gear == "Troll"] <- "Southeast"
# Abalone were only ever caught in southeast
cfec$New.Loc[cfec$New.Loc == "Statewide" & cfec$spec == "ABLN"] <- "Southeast"
```

```
#  Create a temporary (ie., smaller) data frame for calculating with only the necessary fields.
temp2 <- select(cfec,-F_REGION,-F_INPFC,-g_price,-P_TYPE,-P_STATUS,-A_RES,-g_price,-startdt,-landdate,ha


#  Identify the earliest fishing year for each p_holder
startyear <- group_by(temp2,p_holder) %>% summarise(start=min(year))
```

```
#  Join that year to the data
temp2 <- left_join(temp2,startyear)
rm(startyear)


#  Define custom function for calculating the mode
Mode <- function(x) {
     ux <- unique(x)
     ux[which.max(tabulate(match(x, ux)))]
}

#  Calculate initial metrics for each p_holder and year
xx <- group_by(temp2,p_holder,year) %>%
              summarise(StartYear=start[1],
              permits=length(unique(p_serial)),# Number of permits
              permitTypes=length(unique(p_fshy)),# Number of permit types
              landings=length(unique(f_ticket)), # Number of fish tickets (proxy for # of trips)
              boats=length(unique(adfg)),# Number of boats fished by p_holder
              areaSm=length(unique(Location)), # Number of permit areas held (N=51)
              areaLg=length(unique(New.Loc)), # Number of our aggregated permit areas held (N=13)
              statareas=length(unique(stat6)), # Number of different areas from which fish were reporte
              vlength=Mode(VLength)
              ) %>% data.frame

rm(gears,p_fshy,Mode)
```

We now begin to examine Revenues as a function of fishing area and gear for each p_holder and year

Aggregate revenue by area and gear for each p_holder and year

```
# temp.dollars <- group_by(temp2,p_holder,year,Gear,New.Loc) %>%
# summarise(Dollars=sum(g_earn)) for(i in
# 1:length(unique(temp.dollars$Gear))){
# print(ggplot(temp.dollars[temp.dollars$Gear==paste(unique(temp.dollars$Gear)[i]),],aes(x=factor(year)
# + geom_boxplot() + facet_wrap(~New.Loc) +
# ggtitle(paste(unique(temp.dollars$Gear)[i]))) }
```

```
# Tally the dollars of fish by each p_holder during each year for each gear
# type First in long format
tempgear.d <- group_by(temp2, p_holder, year, Gear) %>% summarise(Dollars = sum(g_earn))
# Now in wide format, with zeros filled in.
tempgearwide.d <- dcast(tempgear.d, p_holder + year ~ Gear, value.var = "Dollars")
tempgearwide.d[is.na(tempgearwide.d)] <- 0

# Tally the dollars of fish by each p_holder during each year for each
# permit location First in long format
temparea.d <- group_by(temp2, p_holder, year, New.Loc) %>% summarise(Dollars = sum(g_earn))
# Now in wide format, with zeros filled in.
tempareawide.d <- dcast(temparea.d, p_holder + year ~ New.Loc, value.var = "Dollars")
tempareawide.d[is.na(tempareawide.d)] <- 0
```

```
# Tally the dollars of fish by each p_holder during each year for each
# species First in long format
tempspec.d <- group_by(temp2, p_holder, year, spec) %>% summarise(Dollars = sum(g_earn))
tempspecwide.d <- dcast(tempspec.d, p_holder + year ~ spec, value.var = "Dollars")
# Now in wide format
tempspecwide.d[is.na(tempspecwide.d)] <- 0

# Join each of the above wide format tables. In joining, values of g_earn
# from different tables will become redundant so do not use these dollar
# values.
tempjoin.dollars <- inner_join(xx, tempareawide.d) %>% inner_join(tempgearwide.d) %>%
    inner_join(tempspecwide.d)
# Convert the dollar values to dummy values
tempjoin.dollars[, -c(which(names(tempjoin.dollars) %in% colnames(xx)))] <- lapply(tempjoin.dollars[,
    -c(which(names(tempjoin.dollars) %in% colnames(xx)))], function(x) replace(x,
    x != 0, 1))
# Join the total annual revenue for each p_holder with the above table
revenueMetrics <- group_by(temp2, p_holder, year) %>% summarise(totalRev = round(sum(g_earn)),
    medianRev = round(median(g_earn)), cvRev = round(sd(g_earn)/mean(g_earn))) %>%
    inner_join(tempjoin.dollars) %>% mutate(yearsFishing = (year - StartYear) +
    1) %>% select(p_holder, year, yearsFishing, everything()) %>% data.frame


# For plotting purposes the long format will be more useful
longjoin.dollars <- inner_join(xx, temparea.d) %>% inner_join(tempgear.d) %>%
    inner_join(tempspec.d) %>% mutate(yearsFishing = (year - StartYear) + 1)
```

Now do the same thing but aggregate the pounds landed (instead of the revenue) landed by area, gear,species for each p_holder and year

```
# Tally the pounds of fish by each p_holder during each year for each gear
# type First in long format
tempgear.p <- group_by(temp2, p_holder, year, Gear) %>% summarise(Mass = sum(g_pounds))
# Now in wide format, with zeros filled in.
tempgearwide.p <- dcast(tempgear.p, p_holder + year ~ Gear, value.var = "Mass")
tempgearwide.p[is.na(tempgearwide.p)] <- 0

# Tally the pounds of fish by each p_holder during each year for each permit
# location First in long format
temparea.p <- group_by(temp2, p_holder, year, New.Loc) %>% summarise(Mass = sum(g_pounds))
# Now in wide format, with zeros filled in.
tempareawide.p <- dcast(temparea.p, p_holder + year ~ New.Loc, value.var = "Mass")
tempareawide.p[is.na(tempareawide.p)] <- 0

# Tally the pounds of fish by each p_holder during each year for each
# species First in long format
tempspec.p <- group_by(temp2, p_holder, year, spec) %>% summarise(Mass = sum(g_pounds))
tempspecwide.p <- dcast(tempspec.p, p_holder + year ~ spec, value.var = "Mass")
# Now in wide format
tempspecwide.p[is.na(tempspecwide.p)] <- 0

# For plotting purposes the long format will be more useful
longjoin.pounds <- inner_join(xx, temparea.p) %>% inner_join(tempgear.p) %>%
```

```
    inner_join(tempspec.p) %>% mutate(yearsFishing = (year - StartYear) + 1)

tempjoin.pounds <- inner_join(xx, tempareawide.p) %>% inner_join(tempgearwide.p) %>%
    inner_join(tempspecwide.p)
tempjoin.pounds[, -c(which(names(tempjoin.pounds) %in% colnames(xx)))] <- lapply(tempjoin.pounds[,
    -c(which(names(tempjoin.pounds) %in% colnames(xx)))], function(x) replace(x,
    x != 0, 1))

# Join the total annual revenue for each p_holder with the above table
landedMetrics <- group_by(temp2, p_holder, year) %>% summarise(totalPounds = round(sum(g_pounds)),
    medianPound = round(median(g_pounds)), cvPounds = round(sd(g_pounds)/mean(g_pounds))) %>%
    inner_join(tempjoin.pounds) %>% mutate(yearsFishing = (year - StartYear) +
    1) %>% select(p_holder, year, yearsFishing, everything()) %>% data.frame
```

Combine metrics into a single table

```
AllMetrics.wide <- inner_join(revenueMetrics, landedMetrics) %>% select(p_holder,
    StartYear, yearsFishing, vlength, permits:statareas, totalRev:cvRev, totalPounds:cvPounds,
    everything())
AllMetrics.long <- inner_join(longjoin.pounds, longjoin.dollars) %>% select(p_holder,
    StartYear, yearsFishing, vlength, permits:New.Loc, Gear, spec, Mass, Dollars)

temparea <- inner_join(temparea.p, temparea.d)
tempgear <- inner_join(tempgear.p, tempgear.d)
tempspec <- inner_join(tempspec.p, tempspec.d)

rm(xx, cfec, temp2, revenueMetrics, landedMetrics, longjoin.pounds, longjoin.dollars,
    tempjoin.dollars, tempjoin.pounds, tempspec.p, tempspec.d, temparea.p, temparea.d,
    tempgear.p, tempgear.d)
save.image("cfecModelsetup.RData")
```

Metadata - what do all of the different objects mean?

For modeling, the object most usefull will likely be "AllMetrics.wide"

AllMetrics.long contains combined characteristics for each p_holder and a single column for all species, a single column for all gears, a single column for all areas (New.Loc), a single column for pounds landed, and a single column for dollars earned.

AllMetrics.wide contains combined characteristics for each p_holder and a different column for each species, a different column for each gear, a different column for each areas (New.Loc), a single column total,median and cv of revenue, a single column for total, median, and cv of pounds landed.

temparea contain dollars & pounds for each p_holder by year and by area (New.Loc) - all areas are in a single column

tempareawide.d / *tempareawide.p* contain dollars / pounds for each p_holder by year and by area (New.Loc) - each areas is in a different column

tempgear contain dollars & pounds for each p_holder by year and by gear type - all gears are in a single column

tempgearwide.d / *tempgearwide.p* contain dollars / pounds for each p_holder by year and by gear - each gear is in a different column

tempspec contain dollars & pounds for each p_holder by year and by species - all species are in a single column

tempspecwide.d / *tempspecwide.p* contain dollars / pounds for each p_holder by year and by species - each species is in a different column