

Analyzing portfolio effects in commercial catch data

NCEAS commercial catch portfolio working group

October 30, 2015

Overview

We are interested in multiple metrics of diversity. For this first paper, we plan to divide the analysis into several components:

1. Characterize trends in diversity over time (using GAMs). The metrics of diversity we might want to include are
 - variation in species diversity
 - variation in gear diversity
 - variation in spatial diversity (number of commercial areas fished?)
 - variation in permit holdings over time
2. Using diversity as the predictor, look at absolute or relative change revenue / vessel (or revenue / vessel / days fished) as the response.

For both of the above, there's a large number of other perturbations including EVOS, management changes, climate shifts, changes in gear or processing capacity, hatchery production of salmon, market conditions (such as Japanese crash in mid-1990s), constraints on limited entry or quotas, and biological conditions or shifts (abundance, distribution, size structure, etc).

We'll initially restrict our analysis to data after 1984, since the data from 1975-1984 were dominated by foreign vessels (which 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>

Data processing

Load the data from the saved workspace. This is the raw cfec data that has been merged to have the same columns across years, 1975 - present. The key columns are the species labels ("spec") and species number codes ("specn"), the gear number codes ("gearn"), the vessel id ("adfg"), the area where the catch occurred ("stat6"), the pounds harvested ("g_pounds"), gross earnings ("g_earn").

Note: for any confidentiality concerns, the workspace loaded here has been slightly jittered. This will need to be replaced with the real data for the final model.

```
load("cfec_jitter.Rdata")
# initially only use data 1985+
cfec = cfec[cfec$year > 1984, ]
```

The first data processing step is to aggregate some of the species. This is largely needed because of changes in reporting over time. For example, some of the rockfishes weren't reported to the same species level. For other species, there are unidentified groups, so there's additional aggregation required.

```
# Lump all skates together
cfec$spec = as.character(cfec$spec)
cfec$specn[cfec$specn == 700] = 700
cfec$specn[cfec$specn == 701] = 700
cfec$specn[cfec$specn == 702] = 700
cfec$spec[cfec$specn == 700] = "SKATE"

# Lump all sharks together
cfec$specn[cfec$specn == 691] = 689
cfec$specn[cfec$specn == 689] = 689
cfec$spec[cfec$specn == 689] = "SHARK"

# 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$spec[cfec$specn == 119] = "FLAT.shallow"

# Lump flatfish (deep)
cfec$specn[cfec$specn == 124] = 118 # dover
cfec$specn[cfec$specn == 134] = 118 # greenland turbot
cfec$spec[cfec$specn == 118] = "FLAT.deep"

# Lump rockfish, codes from here:
cfec$specn[cfec$specn == 141] = 144 # POP
cfec$specn[cfec$specn == 136] = 144 # northern
cfec$specn[cfec$specn == 151] = 144 # rougheye
cfec$specn[cfec$specn == 166] = 144 # sharpchin
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 == 137] = 144 # bocaccio
cfec$specn[cfec$specn == 144] = 144 # unid
cfec$spec[cfec$specn == 144] = "ROCK.slope"

cfec$specn[cfec$specn == 138] = 168 # copper
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$spec[cfec$specn == 168] = "ROCK.demersal"

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 == 172] = 169 # dusky
```

```

cfec$specn[cfec$specn == 173] = 169 # dark
cfec$specn[cfec$specn == 142] = 169 # black
cfec$specn[cfec$specn == 167] = 169 # blue
cfec$spec[cfec$specn == 169] = "ROCK.pelagic"

# Lump king crabs
cfec$spec[cfec$spec %in% c("GKCR", "BKCR", "UKCR", "RKCR")] = "KCRB" # king
# Lump tanner crabs
cfec$spec[cfec$spec %in% c("UTCR", "BTCR")] = "TCRB" # tanner
# Lump shrimp - spot prawn, pink shrimp,
cfec$spec[cfec$spec %in% c("SPRW", "PSHP", "USRM")] = "USRM" # shrimp

# drop records with no spp
cfec = cfec[-which(cfec$spec == ""), ]

```

Next, we'll deal with processing the areas. We have multiple types of area codes. The 6-digit codes represent groundfish management areas, the 5-digit codes are state salmon/shellfish areas, and the 3-digit codes are the 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.

```

# Filter out sub-districts and group things at larger areas Lump statistical
# areas by district -- groundfish
cfec$stat6 = as.character(as.numeric(cfec$stat6)) # as.numeric needed because of dumb integer
cfec$areaType = NA
cfec$stat6_char = NA
indx = which(cfec$stat6 > 99999 & cfec$stat6 < 999999)
cfec$stat6_char[indx] = nchar(cfec$stat6[indx])
cfec$areaType[indx] = "groundfish"
# This is just turning the last digit - sub region - into 0
cfec$stat6[indx] = paste(substr(cfec$stat6[indx], 1, 5), "0", sep = "")
cfec$stat6[indx] = as.numeric(cfec$stat6[indx])
# Now do salmon / shellfish - This is just turning the last 2 digits - sub
# region - into 00
indx = which(cfec$stat6 > 9999 & cfec$stat6 < 1e+05)
cfec$stat6[indx] = paste(substr(cfec$stat6[indx], 1, 3), "00", sep = "")
cfec$areaType[indx] = "salmonShellfish"

cfec$stat6 = as.numeric(cfec$stat6)
# Deal with the IPHC areas too. We can verify that nearly all spp caught in
# 3-digit areas are halibut, sort(table(cfec$spec[which(cfec$stat6 <
# 1000)]))

```

Next, we'll aggregate the stat6 codes into larger geographic regions. The focus here is on Prince William Sound, but for contrast we'll include Southeast Alaska, Kodiak W / E, and Cook Inlet.

```

# Add the areas -- 5 digit are salmon/shellfish, 6-digit are groundfish
cfec$area = NA
cfec$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"

# Area to add - Alaska peninsula - Halibut areas 300-340?

```

Exploratory data analysis

Drop all species records that have less than 1000 total observations. This is discarding about 5000 of 20 million records.

```

cfecClean = group_by(cfec, spec) %>% mutate(nspecies = n()) %>% filter(nspecies >
  1000) %>% data.frame

# Also drop out unid groundfish, unid flatfish, smelts, grenadier, english
# sole -- all very rare
cfecClean = cfecClean[-which(cfecClean$spec %in% c("UFLT", "UGRN", "SMLT", "GRDR",
  "EGLS")), ]

```

Drop all species records with gears occurring < 1000 times. This is discarding about 1250 records.

```

cfecClean = group_by(cfecClean, gearn) %>% mutate(ngears = n()) %>% filter(ngears >
  1000) %>% data.frame

```

Delete catch from fisheries that have less than 100 total records. This simplifies the dimension of the fishery codes a lot (from 378 to 234), removes about 3050 records, and doesn't affect the total earnings (these fisheries represent 0.0009 of total g_earn).

```

cfecClean = group_by(cfecClean, p_fshy) %>% mutate(nfshy = n()) %>% filter(nfshy >
  100) %>% data.frame

```

Subset the data to only include vessels that fished at any point in Prince William Sound.

```

cfecPWS = group_by(cfecClean, adfg) %>% mutate(in_PWS = any(area %in% "PWS")) %>%
  as.data.frame() %>% filter(in_PWS) %>% data.frame

```

Alan suggested looking at some coarse plots. First, we can show participation of vessels over time. Halibut fishing pre-1995 needed to get quota by catching halibut in 1988-1990.

```

plot(min(cfecPWS$year):max(cfecPWS$year), specnumber(table(cfecPWS$year, cfecPWS$adfg)),
  xlab = "", ylab = "# Vessels", type = "b")

```

We can also look at species by year

```

# total species
par(mfrow = c(2, 1), mgp = c(2, 1, 0), mai = c(0.5, 0.5, 0.3, 0.1))
plot(min(cfecPWS$year):max(cfecPWS$year), specnumber(table(cfecPWS$year, cfecPWS$spec)),
  xlab = "", ylab = "# Species", type = "b", main = "Catch from vessels that ever fished in PWS")

plot(min(cfecPWS$year):max(cfecPWS$year), specnumber(table(cfecPWS$year[cfecPWS$area ==
  "PWS"], cfecPWS$spec[cfecPWS$area == "PWS"])), xlab = "", ylab = "# Species",
  type = "b", main = "Catch only in PWS")

```

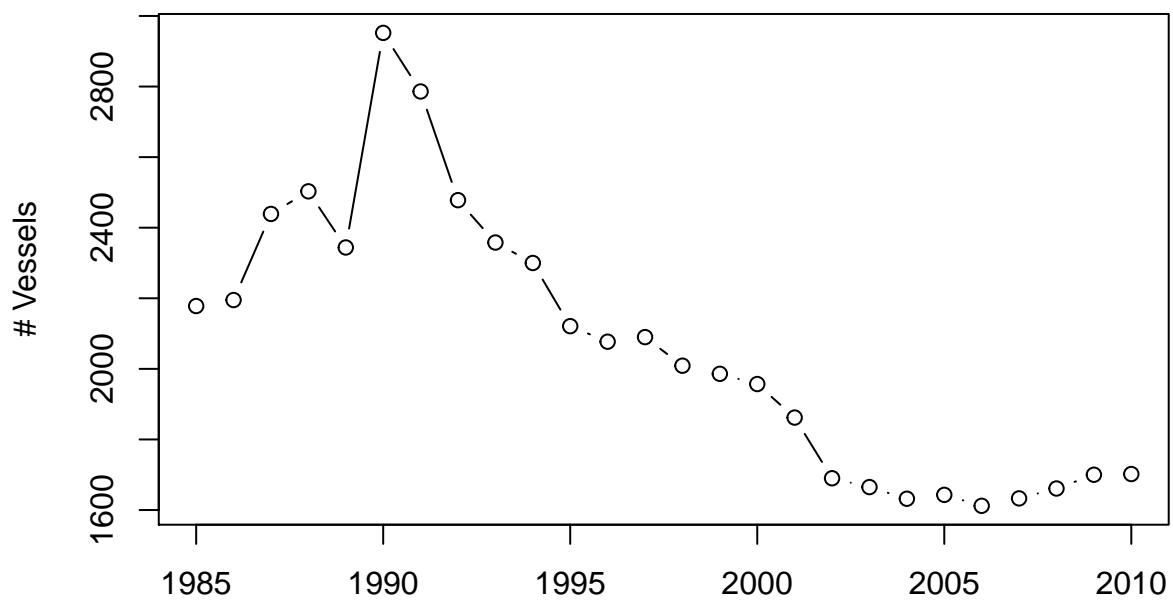


Figure 1: Time series of participation by year

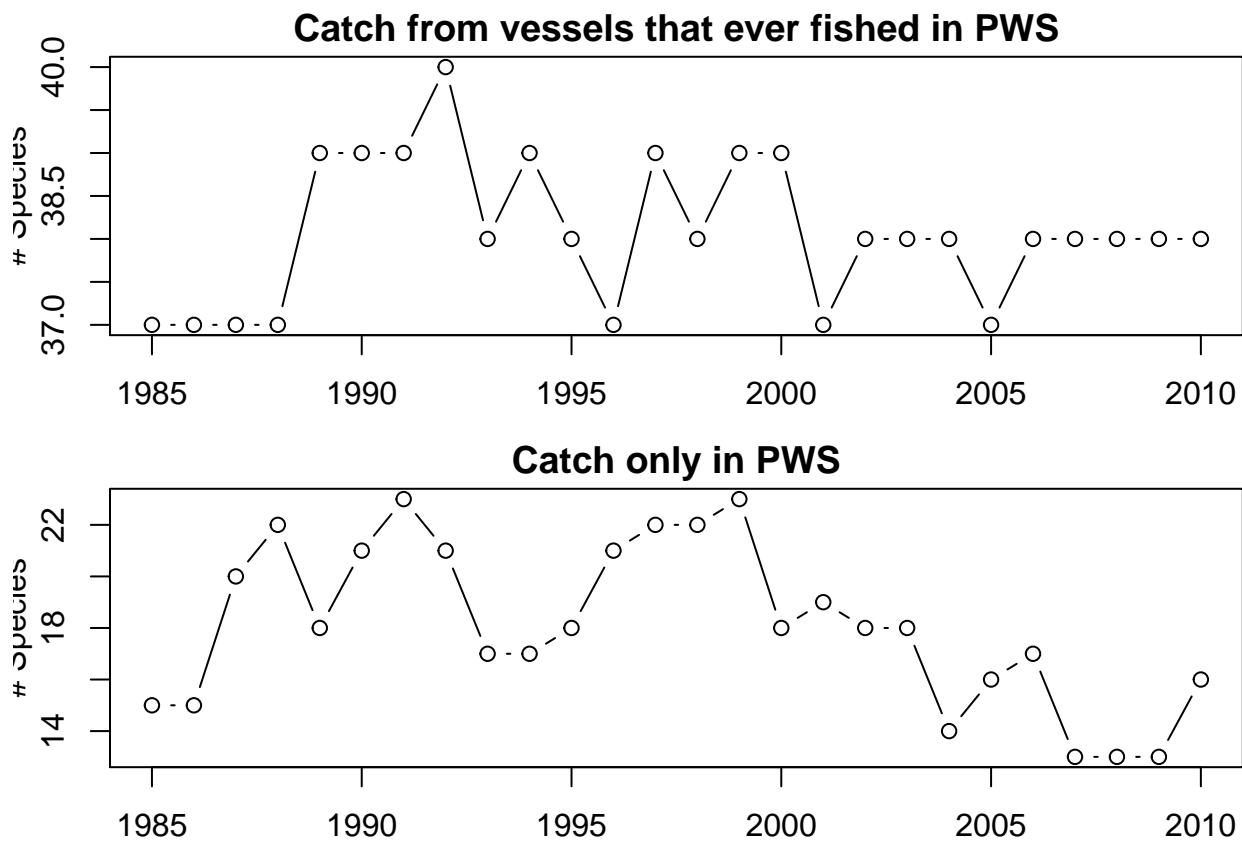
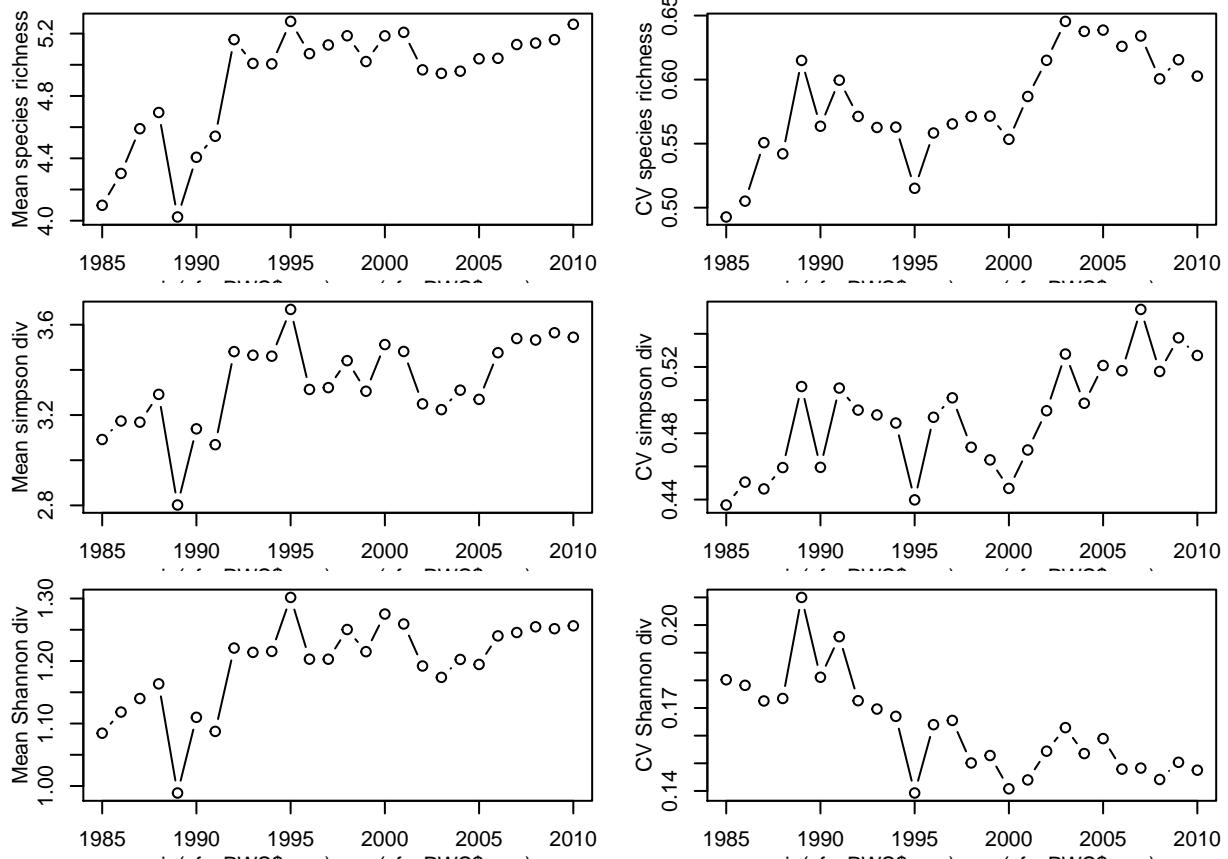
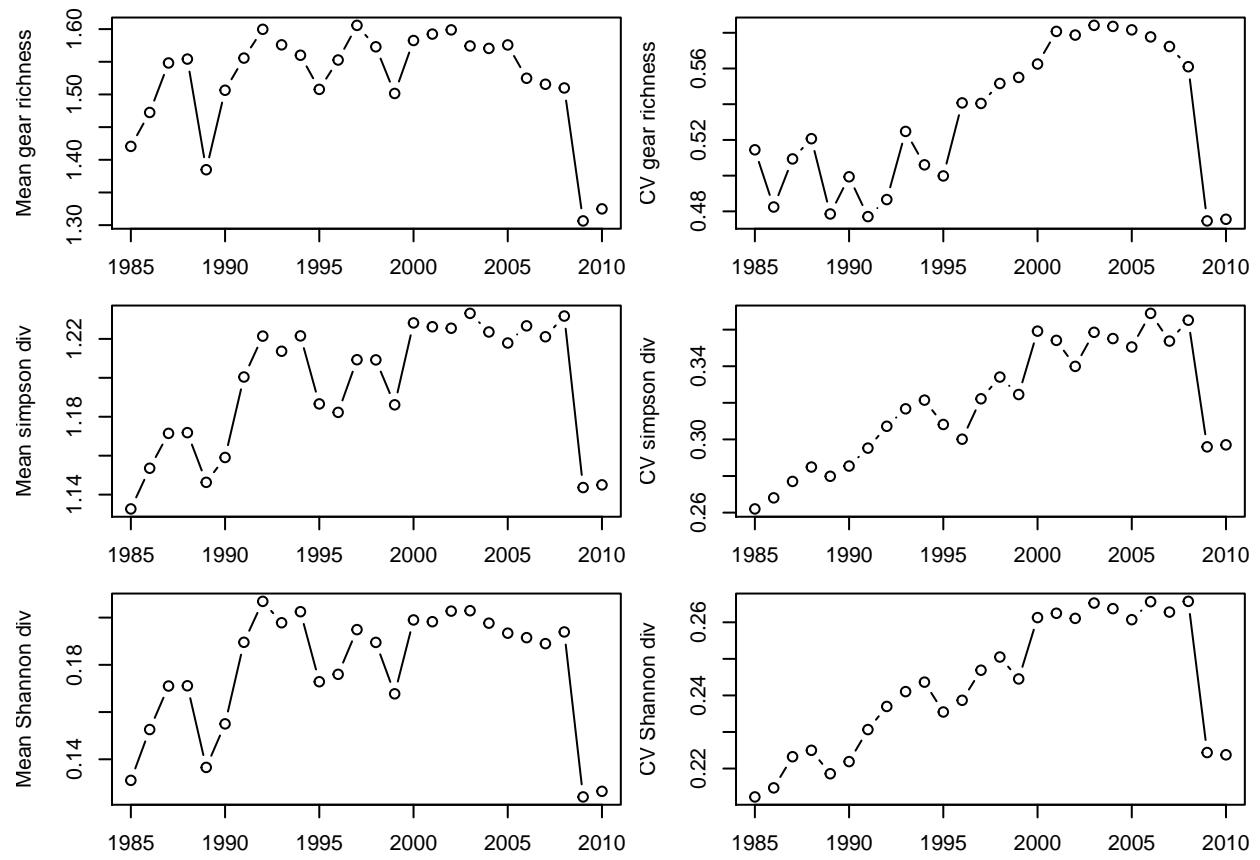


Figure 2: Time series of species caught by year

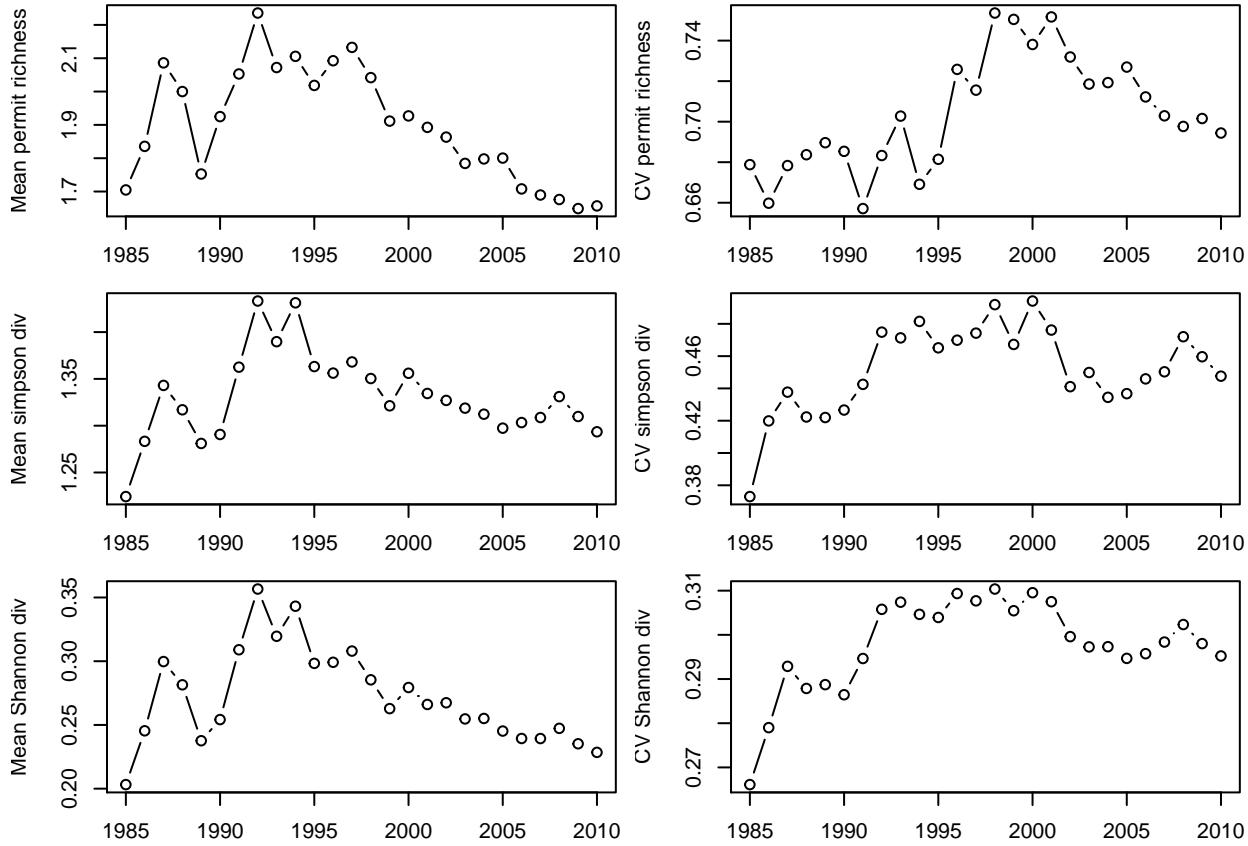
Look at species diversity using Vegan. How has diversity in # species fished changed over time?



We can do similar plots for the number of gears used / vessel



And we can do similar plots for permit types per vessel



Make exploratory plot of vessels that fished only before and after 1989. These aren't shown – but we looked at them in Santa Barbara, and the major take home point is that more vessels left the fishery after 1989 than joined.

Metric of change

We can use absolute or relative earnings over time, by vessel. We adopted the relative measure, because we're comparing vessels of very different size classes. Sizes of individual vessels may change slowly over time. The earning metric is calculated as the first-differenced log earnings from time ($t+1$) to time (t), $\log(\text{earn}_{t+1}) - \log(\text{earn}_t)$.

The key challenge here is that we need a metric of earnings that will allow us to compare different types of vessels (e.g. large, small) that don't overlap in time. For example, we need a metric that standardizes catches or earnings from a small vessel fishing 1985-1990 to that of a large vessel fishing 2000-2005. The relative change is universal across vessels, and can be interpreted as percent change (e.g. a value of -1 = $\exp(-1) = 0.367$, or a ~ 63% decline in earnings between years).

We'll need to be cautious in putting together the covariates if the response is first differenced. If we code up indicators for press / pulse perturbations, those need to be first differenced too.

Aggregating data

The dataset currently is individual records. This needs to be aggregated by vessel / year, so there are totals for each individual vessel. The 3 key dimensions we want to explore with variability are diversity in (1) species, (2) gears, and (3) areas. Diversity in fishery permits is also included for exploration, but permits confounds areas, species, and gears. We need to do some lumping, so we'll initially aggregate

- 10 or more species
- 2 or more areas
- 3 or more gears
- 4 or more fisheries

```
totalPWS = group_by(cfecPWS, adfg, year) %>% summarize(totalearnings = sum(g_earn),
  nspecies = length(unique(spec)), ngear = length(unique(gearn)), narea = length(unique(area)),
  nfishery = length(unique(p_fshy)))
# Lump larger groups
totalPWS$nspecies[totalPWS$nspecies > 10] = 10
totalPWS$narea[totalPWS$narea > 2] = 2
totalPWS$ngear[totalPWS$ngear > 3] = 3
totalPWS$nfishery[totalPWS$nfishery > 4] = 4
```

Add in NAs for missing years. We need these values to be included so that when we first difference, we're not taking differences across > 1 year. When we take the differences, we're also having variables for number of species, gears, etc (along with the differences of each predictor). For the undifferenced data, the number of species, gears, etc. corresponds to the value in the second year. We may want to replace this with a 2-year mean or something.

```
dummyDF = expand.grid(adfg = unique(totalPWS$adfg), year = min(totalPWS$year):max(totalPWS$year),
  stringsAsFactors = F)
# expand the data frame w/missing years
totalPWSExp = left_join(dummyDF, totalPWS) %>% arrange(adfg, year)
totalPWSExp = group_by(totalPWSExp, adfg) %>% mutate(difflog = c(NA, diff(log(totalearnings))),
  diffspec = c(NA, diff(nspecies)), diffgear = c(NA, diff(ngear)), diffarea = c(NA,
  diff(narea)), difffishery = c(NA, diff(nfishery))) %>% filter(!is.na(difflog))
```

Now with our aggregated dataset, we can make histograms of each predictor.

```
par(mfrow = c(2, 2), mai = c(0.6, 0.6, 0.1, 0.1))
# hist(log(totalPWSExp$totalearnings), main='Log of
# earnings', xlab='', col='grey')
hist(totalPWSExp$nspecies, main = "Number of species", xlab = "", col = "grey")
hist(totalPWSExp$ngear, main = "Number of gears", xlab = "", col = "grey")
hist(totalPWSExp$narea, main = "Number of areas", xlab = "", col = "grey")
hist(totalPWSExp$nfishery, main = "Number of fisheries", xlab = "", col = "grey")
```

We can also look at change in these predictors over time – we can also look at the response, which is change in earnings from time t to time t+1.

```
par(mfrow = c(3, 2), mai = c(0.6, 0.6, 0.2, 0.1))
hist(totalPWSExp$difflog, main = "Log of change in earnings", xlab = "", col = "grey")
hist(totalPWSExp$diffspec, main = "Change in species", xlab = "", col = "grey")
hist(totalPWSExp$diffgear, main = "Change in gears", xlab = "", col = "grey")
hist(totalPWSExp$diffarea, main = "Change in areas", xlab = "", col = "grey")
hist(totalPWSExp$difffishery, main = "Change in fisheries", xlab = "", col = "grey")
```

Plotting change in earnings through time

We can start by making plots for individual data points where the dimension doesn't change (e.g. individuals fish 1 species in both time t and time t+1).

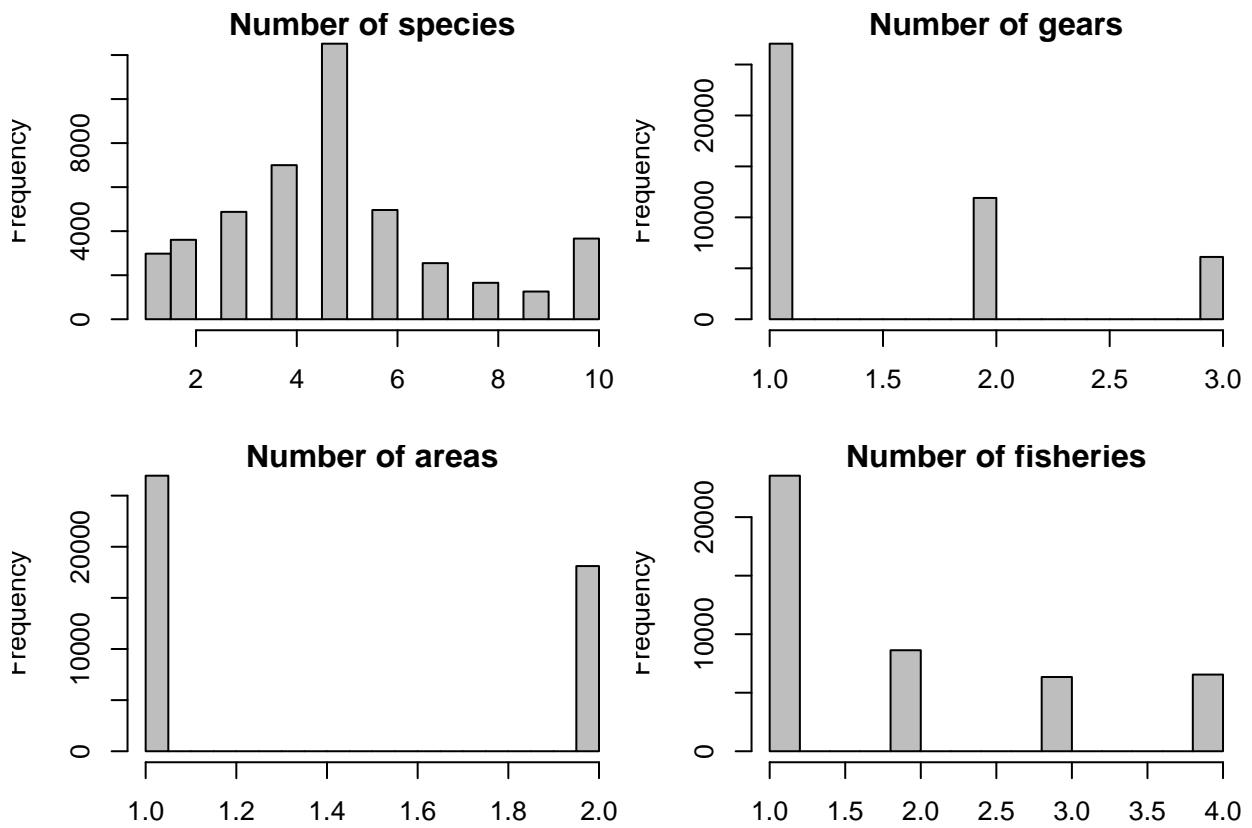


Figure 3: Histogram of predictors (across all years and data)

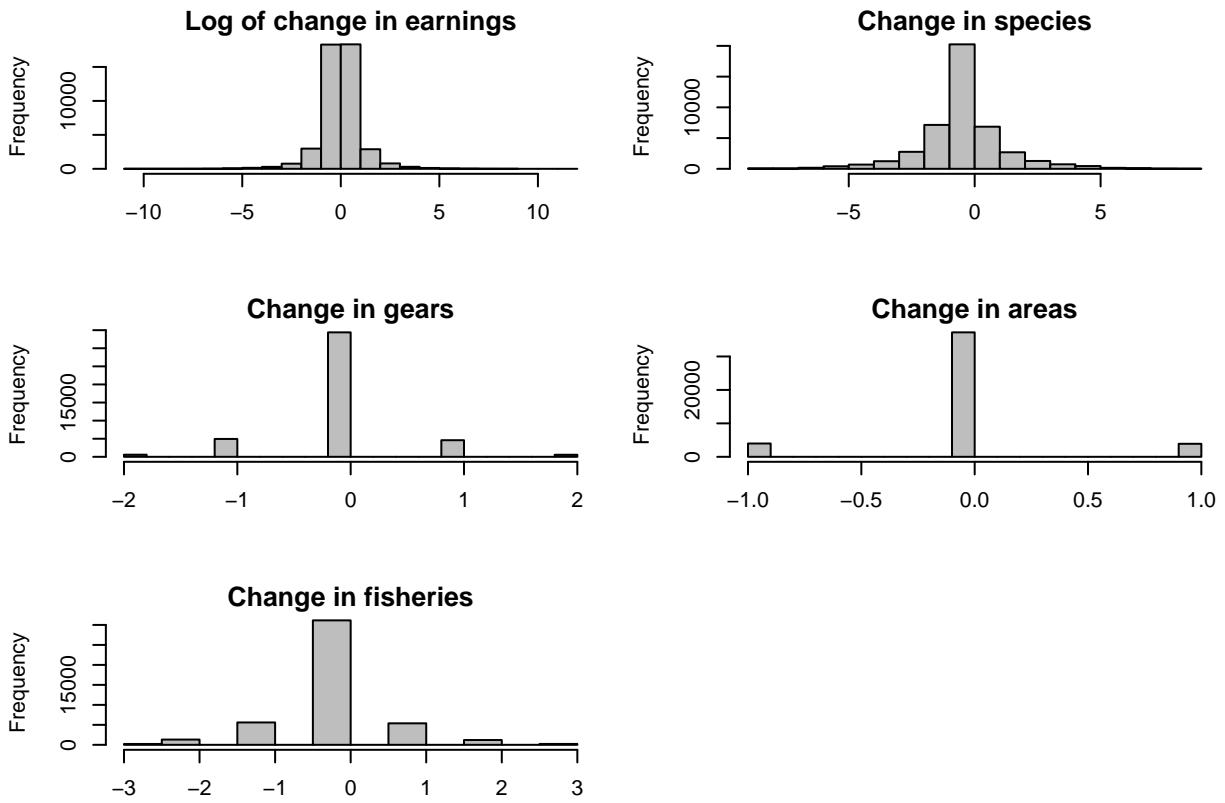


Figure 4: Histogram of change in predictors (across all years and data)

```
group_by(totalPWSEExp[totalPWSEExp$diffspec == 0, ], year, nspecies) %>% summarize(medDiffLog = median(difflog), lower = quantile(difflog, probs = 0.25), upper = quantile(difflog, probs = 0.75)) %>% ggplot(aes(year, medDiffLog)) + geom_line() + facet_wrap(~nspecies) + geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2)

group_by(totalPWSEExp[totalPWSEExp$diffgear == 0, ], year, ngear) %>% summarize(medDiffLog = median(difflog), lower = quantile(difflog, probs = 0.25), upper = quantile(difflog, probs = 0.75)) %>% ggplot(aes(year, medDiffLog)) + geom_line() + facet_wrap(~ngear) + geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2)

group_by(totalPWSEExp[totalPWSEExp$diffarea == 0, ], year, narea) %>% summarize(medDiffLog = median(difflog), lower = quantile(difflog, probs = 0.25), upper = quantile(difflog, probs = 0.75)) %>% ggplot(aes(year, medDiffLog)) + geom_line() + facet_wrap(~narea) + geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2)

group_by(totalPWSEExp[totalPWSEExp$difffishery == 0, ], year, nfishery) %>% summarize(medDiffLog = median(difflog), lower = quantile(difflog, probs = 0.25), upper = quantile(difflog, probs = 0.75)) %>% ggplot(aes(year, medDiffLog)) + geom_line() + facet_wrap(~nfishery) + geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2)
```

We can break the differences out further, by species (across columns) and by number of areas fished. The areas are grouped by vessels fishing in one area (just PWS) versus vessels that fish in PWS and 1 or more other regions. This plot includes some confounding, so don't read too much into it.

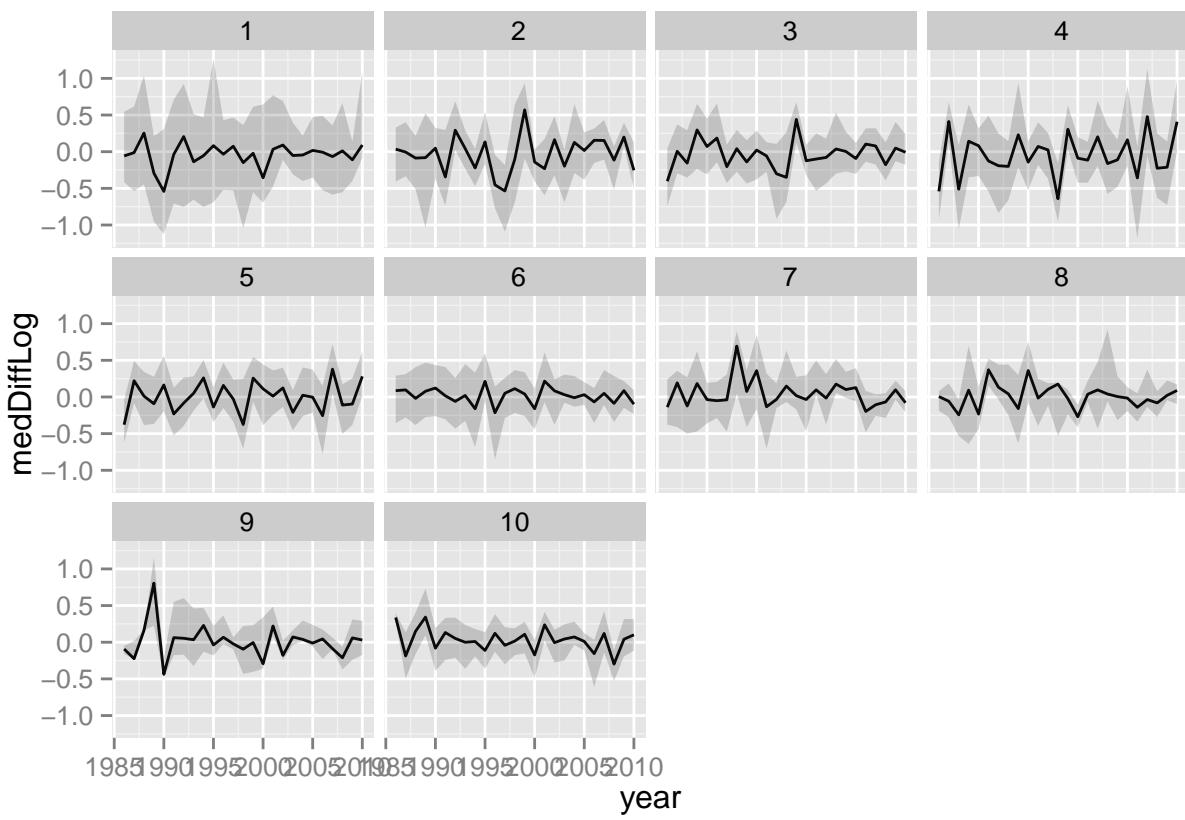


Figure 5: Median and quartiles of change in earnings, by species, for individuals that don't change number of species fished through time

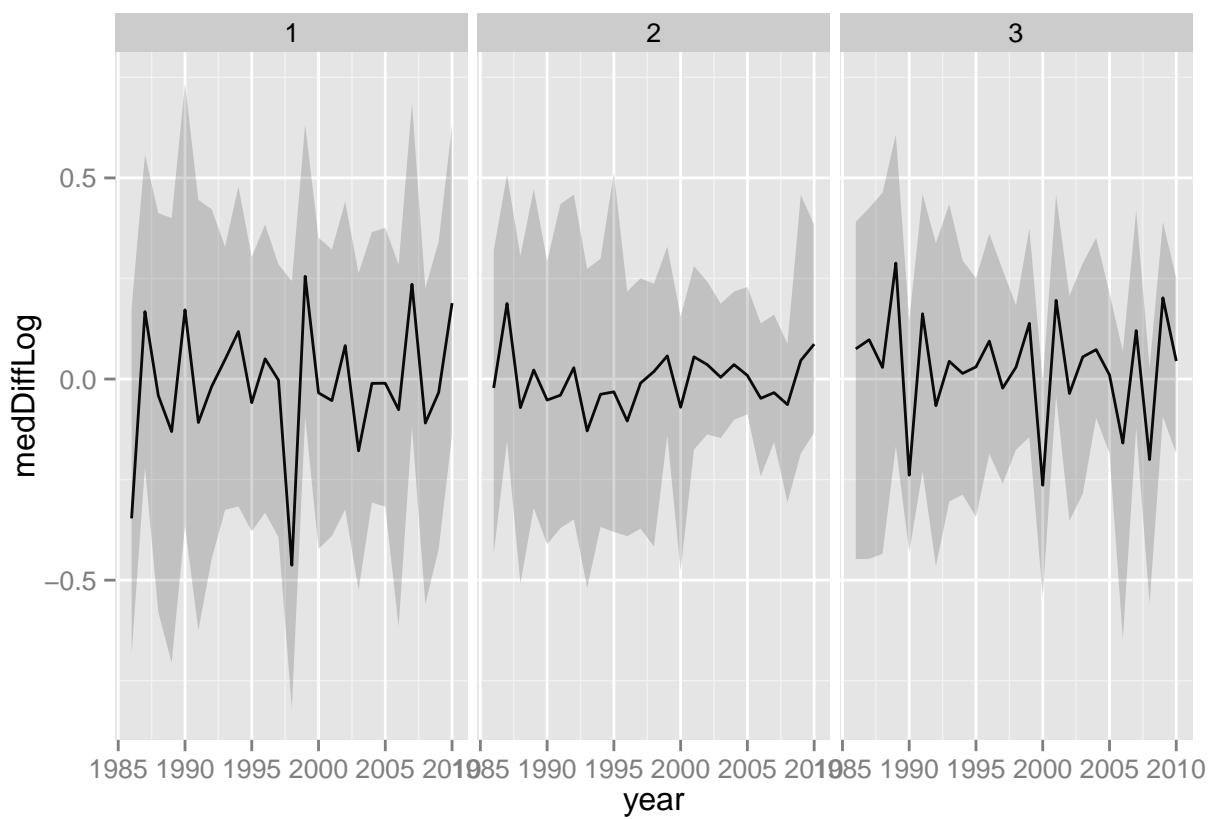


Figure 6: Median and quartiles of change in earnings, by gears, for individuals that don't change number of gears fished through time

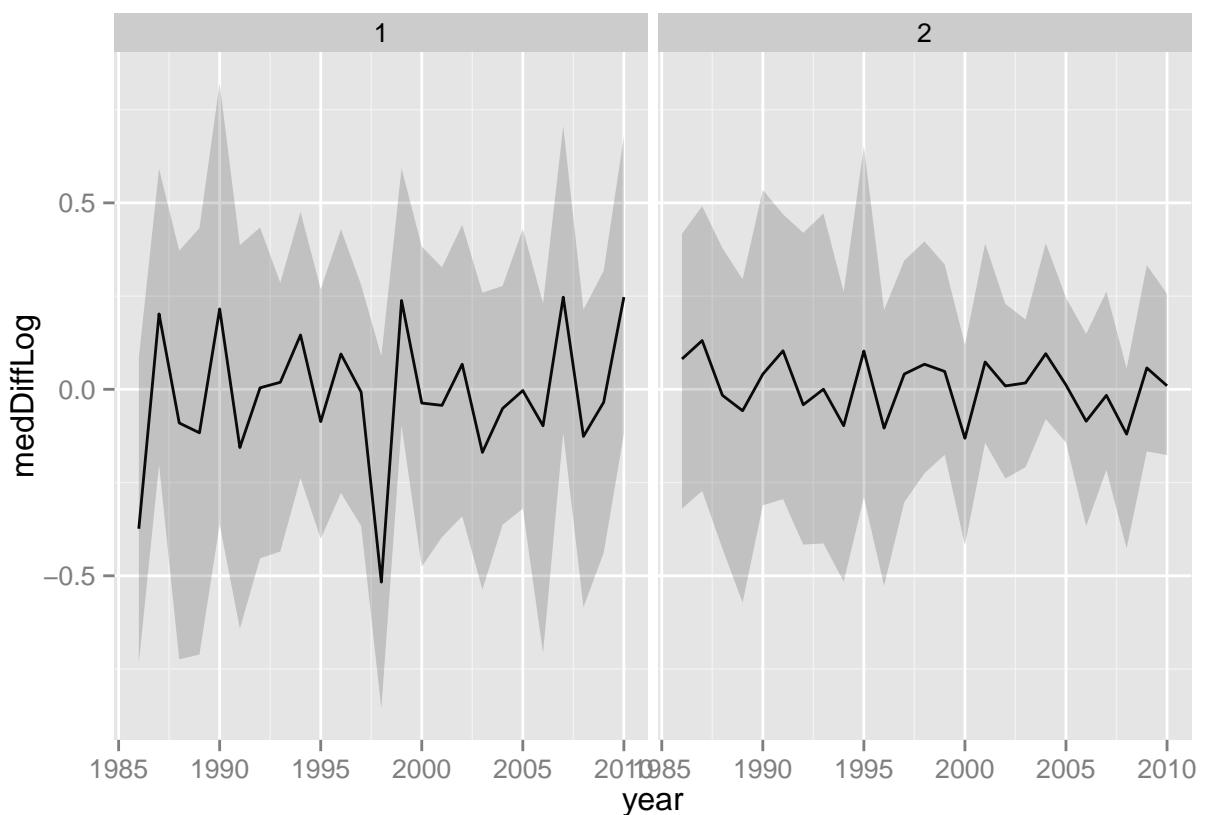


Figure 7: Median and quartiles of change in earnings, by areas, for individuals that don't change number of areas fished through time



Figure 8: Median and quartiles of change in earnings, by areas, for individuals that don't change number of fisheries fished through time

```

totalPWSExp2 = totalPWSExp
totalPWSExp2$nspecies[totalPWSExp2$nspecies > 5] = 5
group_by(totalPWSExp2, year, nspecies, narea) %>% summarize(medDiffLog = median(difflog),
  lower = quantile(difflog, probs = 0.25), upper = quantile(difflog, probs = 0.75)) %>%
  ggplot(aes(year, medDiffLog)) + geom_line() + facet_grid(narea ~ nspecies) +
  geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2)

```

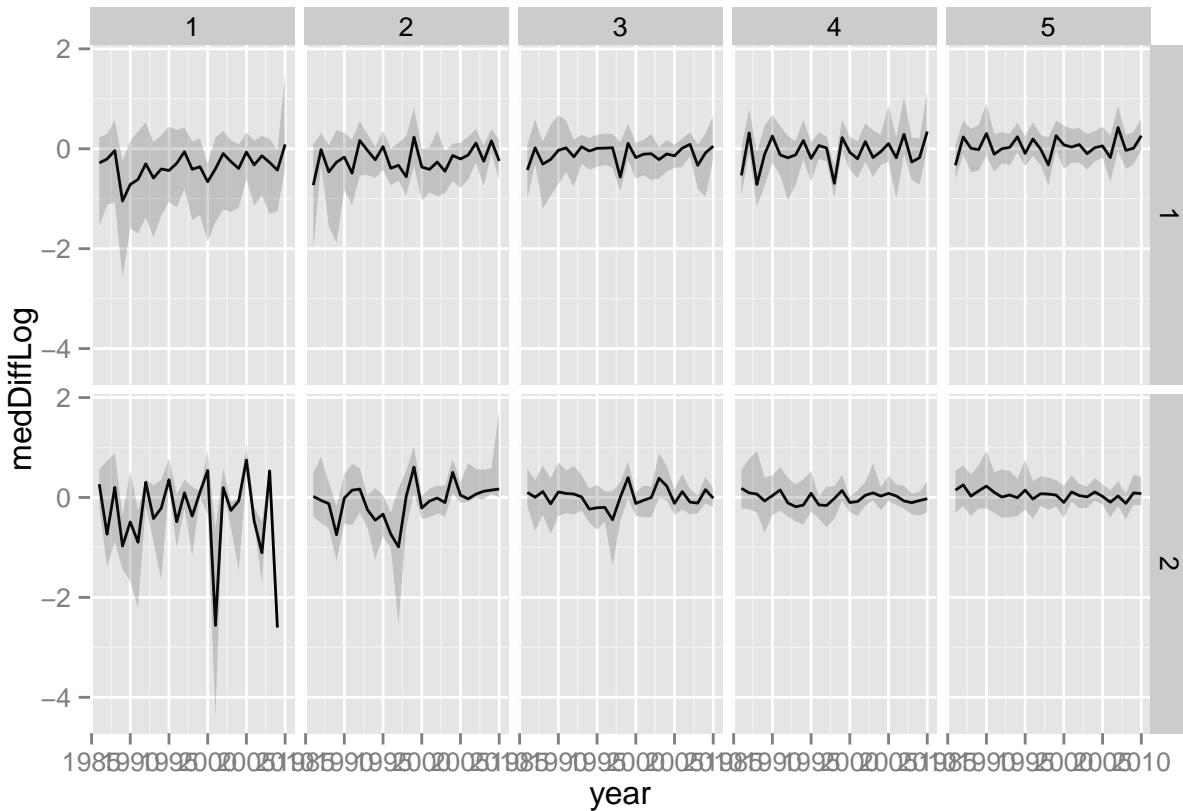


Figure 9: Median and quartiles of change in earnings, by species x areas

Make some summary tables of vessels fishing in 1988 / 1989 that only landed one species. Make tables of the most common species landed, both by records and earnings.

```

vOne = filter(totalPWSExp, nspecies==1 & diffspec==0) %>%
  select(adfg) #>% #select(year,adfg)
  unique

LJ = left_join(vOne, cfecPWS)

# Look at species that fishers are specializing on
head(table(LJ$spec[LJ$year%in%seq(1988,1989)])) %>%
  sort %>%
  rev

##  

## SOCK CHUM PINK COHO PHLB SABL  

## 3469 3337 3054 2134 2102 2093

```

```

# Look at earnings of fishers that people are specializing on
LJ89 = LJ[LJ$year==1989,]
LJ89table = aggregate(g_earn ~ spec, LJ89, sum)
head(LJ89table[with(LJ89table, order(-g_earn)), ])

```

```

##      spec    g_earn
## 17 PLCK 85297234
## 16 PINK 8160776
## 15 PHRG 6341951
## 12 OTCR 5660752
## 23 SABL 3968477
## 14 PHLB 3767853

```

We can also get really into the weeds and make plots of individual annual landings, by vessel and species. For confidentiality, these aren't shown but the code can be run.

Fitting models to data

Before we fit anything, we need to calculate the mean of each predictor (# species, gears, areas) over the 2 year period. This will basically scale the effect, and will be included alongside predictors representing the change in each of these variables.

Are a priori expectation is that if diversity (species, gear, area) is beneficial, all of the differenced predictors (diffspec, diffgear, diffarea) will have positive effects on the response (change in earnings). But if the differenced values are small, it means individual strategy is relatively constant, and we'd also expect the effects of diversification on the mean values (nspecies, narea, ngear).

```

totalPWSEExp$nspecies = (totalPWSEExp$nspecies + (totalPWSEExp$nspecies + totalPWSEExp$diffspec))/2
totalPWSEExp$ngear = (totalPWSEExp$ngear + (totalPWSEExp$ngear + totalPWSEExp$diffgear))/2
totalPWSEExp$narea = (totalPWSEExp$narea + (totalPWSEExp$narea + totalPWSEExp$diffarea))/2
totalPWSEExp$narea = (totalPWSEExp$narea + (totalPWSEExp$narea + totalPWSEExp$diffarea))/2

```

Start with simple linear model, no random effects

```

lm.simple = lm(difflog ~ as.factor(year) + nspecies + ngear + narea + diffspec +
  diffgear + diffarea, data = totalPWSEExp)
summary(lm.simple)

```

```

##
## Call:
## lm(formula = difflog ~ as.factor(year) + nspecies + ngear + narea +
##     diffspec + diffgear + diffarea, data = totalPWSEExp)
##
## Residuals:
##      Min        1Q        Median       3Q        Max 
## -10.0268   -0.4301   -0.0004    0.4269   10.4076 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.222461   0.026000  -8.556 < 2e-16 ***
## as.factor(year)1987  0.266405   0.030593   8.708 < 2e-16 ***

```

```

## as.factor(year)1988 0.114734 0.030230 3.795 0.000148 ***
## as.factor(year)1989 0.143913 0.030860 4.663 3.12e-06 ***
## as.factor(year)1990 0.289167 0.030435 9.501 < 2e-16 ***
## as.factor(year)1991 0.148723 0.029466 5.047 4.50e-07 ***
## as.factor(year)1992 0.131464 0.029526 4.452 8.51e-06 ***
## as.factor(year)1993 0.196324 0.029956 6.554 5.67e-11 ***
## as.factor(year)1994 0.271570 0.030083 9.027 < 2e-16 ***
## as.factor(year)1995 0.259856 0.030473 8.527 < 2e-16 ***
## as.factor(year)1996 0.207447 0.030589 6.782 1.20e-11 ***
## as.factor(year)1997 0.121025 0.030655 3.948 7.89e-05 ***
## as.factor(year)1998 0.013660 0.030762 0.444 0.656999
## as.factor(year)1999 0.388037 0.030894 12.560 < 2e-16 ***
## as.factor(year)2000 0.062854 0.031015 2.027 0.042710 *
## as.factor(year)2001 0.187428 0.031306 5.987 2.15e-09 ***
## as.factor(year)2002 0.249189 0.032002 7.787 7.03e-15 ***
## as.factor(year)2003 0.122102 0.032360 3.773 0.000161 ***
## as.factor(year)2004 0.216806 0.032506 6.670 2.59e-11 ***
## as.factor(year)2005 0.254915 0.032530 7.836 4.74e-15 ***
## as.factor(year)2006 0.007937 0.032571 0.244 0.807491
## as.factor(year)2007 0.405119 0.032511 12.461 < 2e-16 ***
## as.factor(year)2008 0.017982 0.032438 0.554 0.579333
## as.factor(year)2009 0.177494 0.032339 5.488 4.08e-08 ***
## as.factor(year)2010 0.455726 0.032219 14.145 < 2e-16 ***
## nspecies 0.007672 0.002678 2.865 0.004178 **
## ngear -0.019957 0.009369 -2.130 0.033167 *
## narea 0.017503 0.012105 1.446 0.148203
## diffspec 0.227038 0.004003 56.719 < 2e-16 ***
## diffgear 0.019480 0.012958 1.503 0.132745
## diffarea 0.126449 0.021337 5.926 3.12e-09 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9289 on 45037 degrees of freedom
## Multiple R-squared: 0.1828, Adjusted R-squared: 0.1822
## F-statistic: 335.7 on 30 and 45037 DF, p-value: < 2.2e-16

```

We can also fit a gam, with spline over year, and basically get the same result. These models show that changing strategy (increasing gears used or areas fished correlates with more earnings, but the overall strategy of more areas or gears doesn't have a positive effect). In contrast, the effect of species is positive in both, suggesting that fishing more species translates to positive revenue. Other splines can be included (nspecies, diffspec) but the effects are basically linear (not shown).

```

gam.simple = gam(difflog ~ s(year) + nspecies + ngear + narea + diffspec + diffgear +
diffarea, data = totalPWSEExp)
summary(gam.simple)

```

```

##
## Family: gaussian
## Link function: identity
##
## Formula:
## difflog ~ s(year) + nspecies + ngear + narea + diffspec + diffgear +
##     diffarea

```

```

## 
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.032724  0.015259 -2.145  0.03200 *  
## nspecies     0.007532  0.002693  2.797  0.00516 ** 
## ngear        -0.021982  0.009417 -2.334  0.01959 *  
## narea         0.018484  0.012172  1.519  0.12887    
## diffspec     0.228529  0.004007 57.026 < 2e-16 ***  
## diffgear     0.024830  0.012999  1.910  0.05613 .  
## diffarea     0.130315  0.021449  6.075 1.25e-09 ***  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Approximate significance of smooth terms:
##          edf Ref.df      F p-value    
## s(year) 8.88 8.996 19.39 <2e-16 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## R-sq.(adj) = 0.173 Deviance explained = 17.3%
## GCV = 0.87302 Scale est. = 0.87272 n = 45068

```

Extending model across areas

First, it's useful to look at the number of vessels that can be uniquely assigned to each of our areas. This shows that > 80% of vessels never fish in more than 1 of our coarse areas (Prince William Sound, Cook Inlet, Bristol Bay, E/W Kodiak, etc), so we can start our analysis focusing on those.

```

tableVesselArea = table(cfecClean$adfg, cfecClean$area)
tableVesselArea = ceiling(tableVesselArea/1e+10)
T = table(apply(tableVesselArea, 1, sum))
m = matrix(NA, 3, 2)
colnames(m) = c("# Vessels", "% Total")
rownames(m) = c("1", "2", "3+")
m[, 1] = c(T[2], T[3], sum(T[-c(1:3)]))
m[, 2] = round(m[, 1]/sum(m[, 1]), 3)

kable(m) # Make simple table w/ knitr

```

	# Vessels	% Total
1	18569	0.813
2	2881	0.126
3+	1404	0.061

Next, we'll extract the vessels that are only assigned to one area, subsetting the entire dataset to just these vessels.

```

singleAreaVessels = rownames(tableVesselArea)[which(apply(tableVesselArea, 1,
sum) == 1)]
cfecClean$adfg = as.character(cfecClean$adfg)

```

```
cfecSingles = group_by(cfecClean, adfg) %>% mutate(is_Single = any(adfg %in%
  singleAreaVessels)) %>% as.data.frame() %>% filter(is_Single) %>% data.frame
```

We can re-run the code above to total the response / predictors by calendar year. We'll add a column for area.

```
totalSingles = group_by(cfecSingles, adfg, year) %>% summarize(totalearnings = sum(g_earn),
  nspecies = length(unique(spec)), ngear = length(unique(gearn)), narea = length(unique(area)),
  nfishery = length(unique(p_fshy)))
# Lump larger groups
totalSingles$nspecies[totalSingles$nspecies > 10] = 10
totalSingles$narea[totalSingles$narea > 2] = 2
totalSingles$ngear[totalSingles$ngear > 3] = 3
totalSingles$nfishery[totalSingles$nfishery > 4] = 4

# Add areas
totalSingles$area = NA
for (i in 1:length(colnames(tableVesselArea))) {
  totalSingles$area[totalSingles$adfg %in% rownames(tableVesselArea)[which(tableVesselArea[, i] == 1)]] = colnames(tableVesselArea)[i]
}

# Fill in missing years and difference
dummyDF = expand.grid(adfg = unique(totalSingles$adfg), year = min(totalSingles$year):max(totalSingles$year),
  stringsAsFactors = F)
# expand the data frame w/missing years
totalSingles = left_join(dummyDF, totalSingles) %>% arrange(adfg, year)
totalSingles = group_by(totalSingles, adfg) %>% mutate(difflog = c(NA, diff(log(totalearnings))),
  diffspec = c(NA, diff(nspecies)), diffgear = c(NA, diff(ngear)), diffarea = c(NA,
  diff(narea)), difffishery = c(NA, diff(nfishery))) %>% filter(!is.na(difflog))
```

Fit another gam, with shared coefficients across areas, but different splines over time by area (Note ‘narea’ and ‘diffarea’ aren’t included because this model only includes vessels fishing in one area). These don’t include quartiles like previous plots, but show that annual patterns are different (and more variable, as function of sample size) across regions. We’ll use something like hierarchical linear models to do a better job of modeling these parameters.

```
gam.area = gam(difflog ~ area + s(year, by = as.factor(area)) + area * nspecies +
  area * diffspec + ngear + diffgear, data = totalSingles)
summary(gam.area)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## difflog ~ area + s(year, by = as.factor(area)) + area * nspecies +
##       area * diffspec + ngear + diffgear
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -0.003787   0.014166  -0.267  0.78919
```

```

## areaCookInlet      -0.041818  0.034578 -1.209  0.22651
## areaKodiak.east   -0.014197  0.042388 -0.335  0.73769
## areaKodiak.west   -0.479638  0.147314 -3.256  0.00113 ** 
## areaPWS            -0.043731  0.027672 -1.580  0.11404
## areaSoutheast     -0.048761  0.016597 -2.938  0.00330 ** 
## nspecies           0.002925  0.004289  0.682  0.49531
## diffspec           0.108782  0.005118 21.254 < 2e-16 *** 
## ngear              -0.007828  0.006434 -1.217  0.22377
## diffgear            0.099060  0.007330 13.514 < 2e-16 *** 
## areaCookInlet:nspecies 0.001514  0.008447  0.179  0.85773
## areaKodiak.east:nspecies 0.001559  0.008073  0.193  0.84692
## areaKodiak.west:nspecies 0.074835  0.030688  2.439  0.01475 * 
## areaPWS:nspecies     0.008386  0.006830  1.228  0.21952
## areaSoutheast:nspecies 0.004720  0.004493  1.050  0.29354
## areaCookInlet:diffspec 0.345489  0.010282 33.600 < 2e-16 *** 
## areaKodiak.east:diffspec 0.119384  0.012674  9.420 < 2e-16 *** 
## areaKodiak.west:diffspec 0.414794  0.047631  8.709 < 2e-16 *** 
## areaPWS:diffspec      0.230172  0.008325 27.649 < 2e-16 *** 
## areaSoutheast:diffspec 0.125694  0.005668 22.178 < 2e-16 *** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                               edf Ref.df      F p-value
## s(year):as.factor(area)BristolBay 8.997  9.000 129.380 < 2e-16 ***
## s(year):as.factor(area)CookInlet   8.905  8.997 35.330 < 2e-16 ***
## s(year):as.factor(area)Kodiak.east 8.433  8.908  2.977 0.00161 ** 
## s(year):as.factor(area)Kodiak.west 3.909  4.813  3.747 0.00263 ** 
## s(year):as.factor(area)PWS        8.889  8.996 22.667 < 2e-16 *** 
## s(year):as.factor(area)Southeast  7.174  8.204 12.344 < 2e-16 *** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.167  Deviance explained = 16.7%
## GCV = 0.8315  Scale est. = 0.83096  n = 103653

par(mfrow = c(3, 2), mgp = c(2, 1, 0), mai = c(0.5, 0.5, 0.3, 0.1))
plot(gam.area)

```

To do list

1. Fit GAMs to diversity metrics to characterize change over time. This could be done with the full dataset, 1975-present.
2. We're going to repeat this kind of analysis on different areas. Some things to consider are:
 - for estimation of random effects, we may need to only include vessels with 3+ records
 - we may want to think about avoiding double counting. The way we have the sorting set up now, a vessel fishing in SEAK and PWS would be counted in both areas / analyses

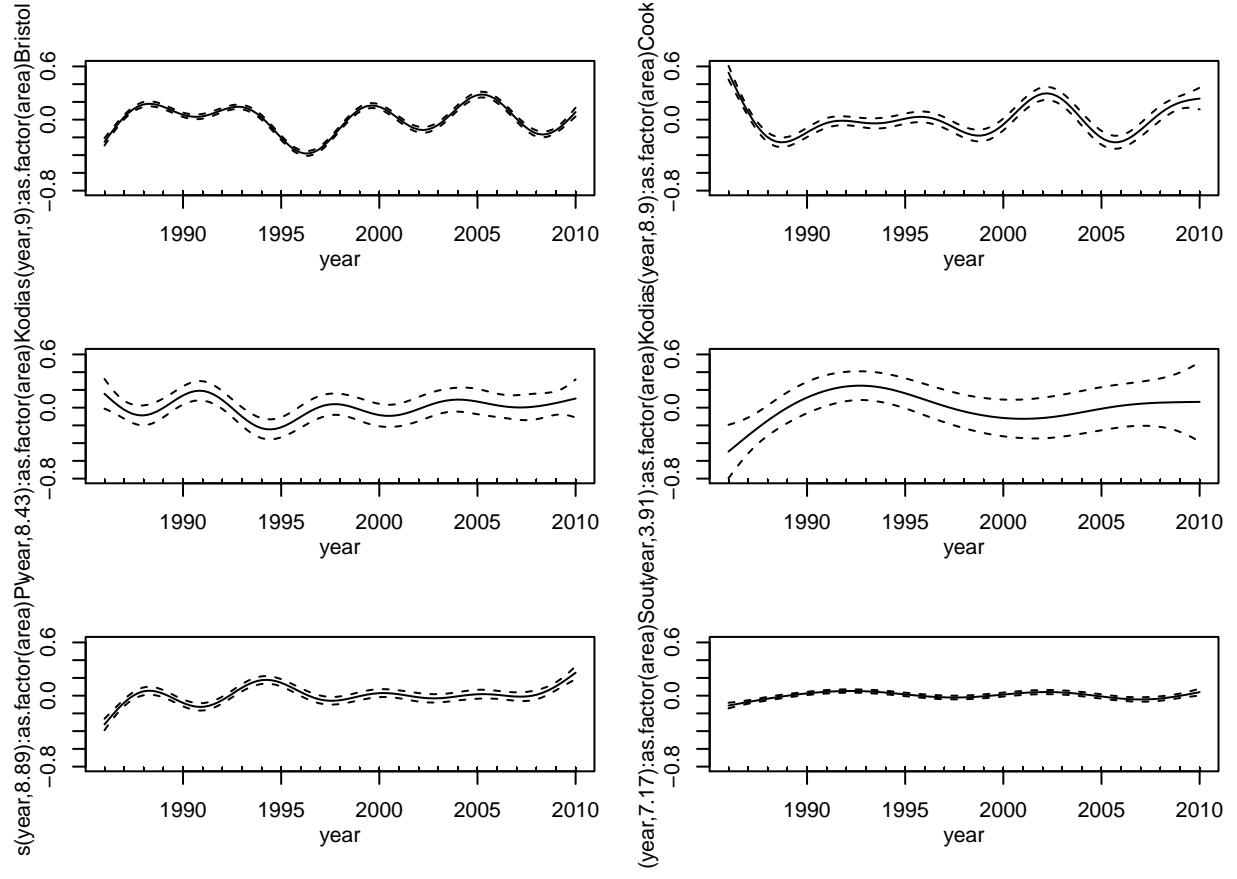


Figure 10: Simple GAM model, grouped by area, with area specific effects of species and change in species fished. Gears, change in gears, areas, and change in areas fished also included. Plots show mean smooth year term, with SE.

3. For comparison between areas, we're interested in characterizing differences in year to year patterns, whether or not diversity relationships/coefficients are shared or random across areas (shared vs random effects), and possibly whether impacts of diversity are changing over time (think DLMs).

Data issues:

1. Why are there no data for longline (gearn == 6) in 2009-2010?