

7_Vis_CCM_Sig_Causal_Vars

Kurt Ingeman

6/27/2021

```
rm(list = ls())

library(here)

## here() starts at /Users/kurtingeman/github/CORE
library(tidyr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(RColorBrewer)
library(ggplot2)

df <- read.csv(here("CORE EDM Visual", "CORE_reduce_CCM_vars", "CCM_twin_95.csv"))
# 250 vars

# ensure that max will work when all values negative
# df[is.na(df)] <- -99

# split df into the rec4 and rec5 values
df4 <- df %>%
  select(1:8) %>%
  rename(
    ESU = ESU_rec4n,
    IMN = Imnaha_rec4n,
    MFS = Middle.Fork.Salmon_rec4n,
    UPS = Upper.Salmon_rec4n
  )

df5 <- df %>%
  select(1:4, 9:12) %>%
  rename(
    ESU = ESU_rec5n,
    IMN = Imnaha_rec5n,
    MFS = Middle.Fork.Salmon_rec5n,
```

```
UPS = Upper.Salmon_rec5n
)
```

1. Are predators drivers of salmon recruitment

evidence: Exceed CCM thresholds, can recover causal vars from salmon dynamics

Rec4: Visualize the distribution of uRho values for broadest categories of causal variables

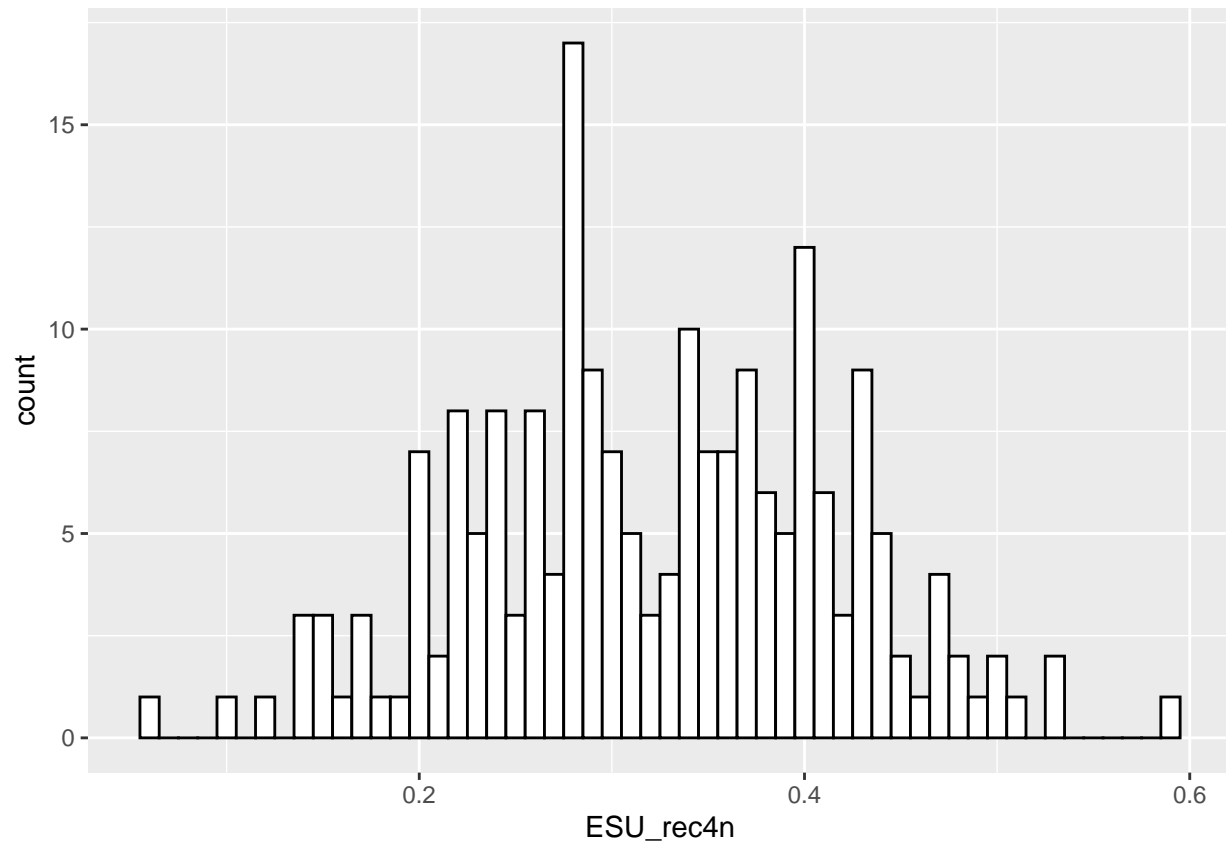
Create new variables to distinguish oceanographic, biological, and human variables

```
df <- df %>%
  filter(!cat == "flow") %>%
  filter(!cat == "hseal") %>%
  mutate(domain = case_when(
    cat == "arc" | cat == "pdo" | cat == "upw" | cat == "npgo" ~ "ocean",
    cat == "ssl" | cat == "csl" | cat == "orca" | cat == "hseal" ~ "predator",
    cat == "harv" | cat == "hatch" ~ "human")) %>%
  group_by(domain) %>%
  mutate(mu_ESU = mean(ESU_rec4n, na.rm=TRUE)) %>%
  ungroup()
```

Histogram of distribution of uRho values, facet by domain

```
ggplot(df, aes(x=ESU_rec4n)) + geom_histogram(color="black", fill="white", binwidth=0.01)
```

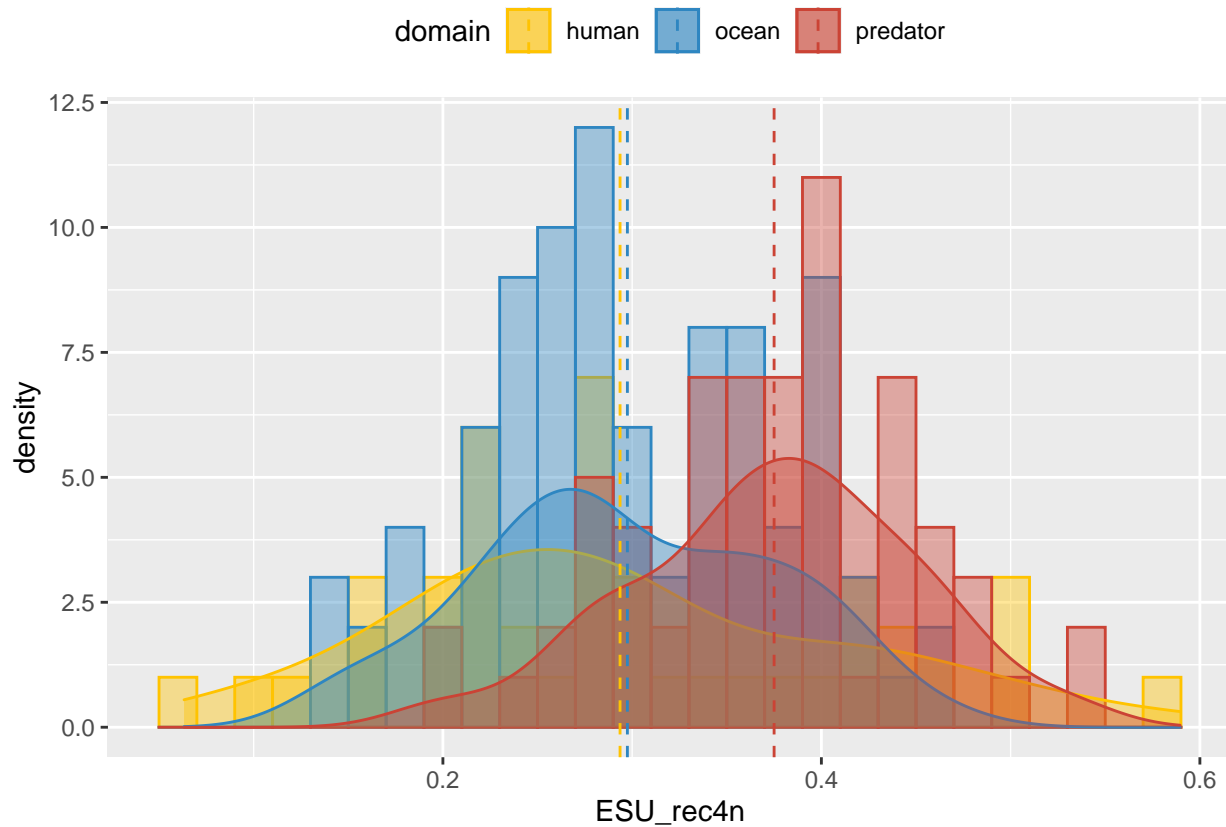
```
## Warning: Removed 22 rows containing non-finite values (stat_bin).
```



```
ggplot(df, aes(x = ESU_rec4n, color = domain, fill = domain )) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.4) +
  geom_density(alpha = 0.4) +
  geom_vline(aes(xintercept = mu_ESU, color = domain),
             linetype = "dashed") +
  theme(legend.position = "top") +
  scale_color_manual(values=c("#FFC300", "#2E86C1", "#CB4335")) +
  scale_fill_manual(values=c("#FFC300", "#2E86C1", "#CB4335"))
```

```
## Warning: Removed 22 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 22 rows containing non-finite values (stat_density).
```

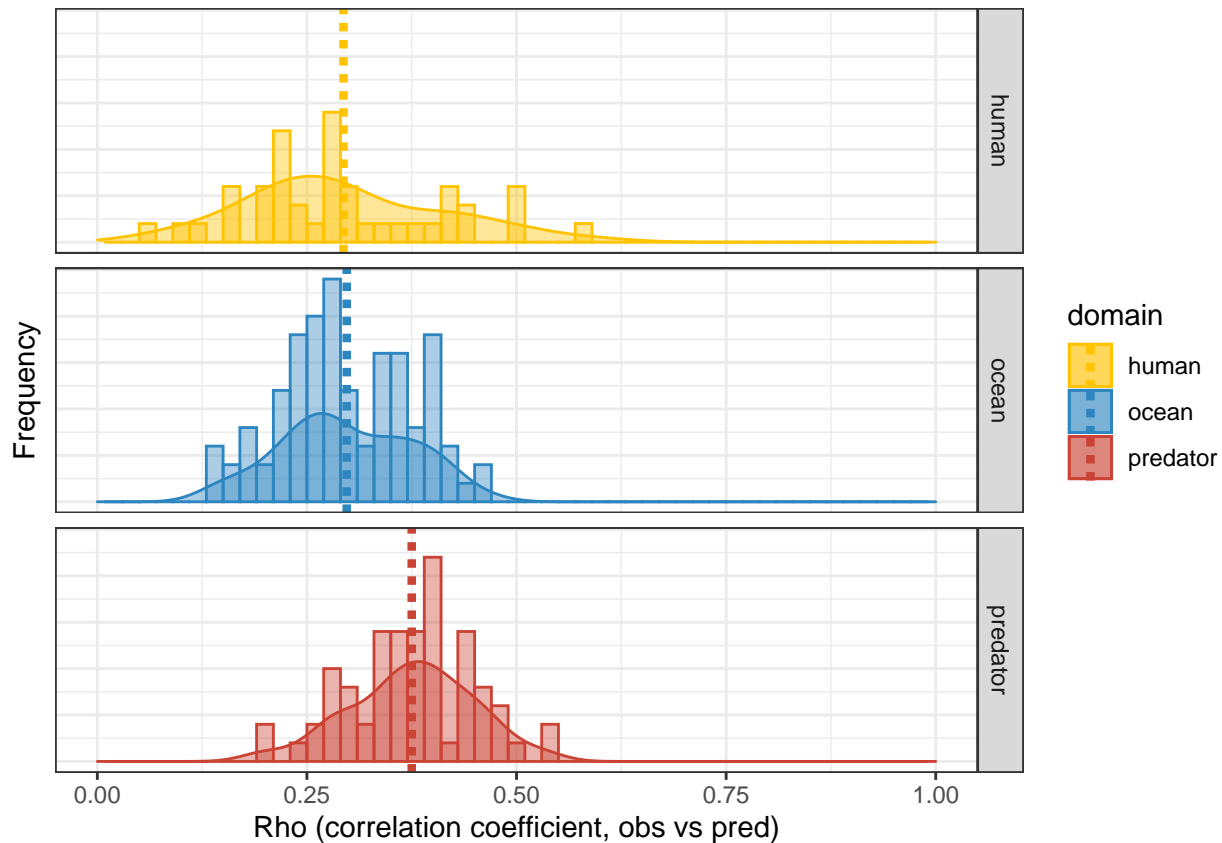


```
ggplot(df, aes(x = ESU_rec4n, color = domain, fill = domain )) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.4) +
  geom_density(alpha = 0.4) +
  scale_color_manual(values=c("#FFC300", "#2E86C1", "#CB4335")) +
  scale_fill_manual(values=c("#FFC300", "#2E86C1", "#CB4335")) +
  facet_grid(domain ~ .) +
  geom_vline(aes(xintercept = mu_ESU, color = domain),
             lty = "11", lwd = 1.5) +
  theme_bw() +
  theme(axis.ticks.y = element_blank(),
        axis.text.y = element_blank()) +
  ylab("Frequency") + xlab("Rho (correlation coefficient, obs vs pred)") +
  xlim(c(0, 1))
```

Warning: Removed 22 rows containing non-finite values (stat_bin).

Warning: Removed 22 rows containing non-finite values (stat_density).

Warning: Removed 6 rows containing missing values (geom_bar).



Are predators important drivers of salmon recruitment? Yes, numerous predator time series exceed CCM thresholds of $\rho > 0$ and increasing with library size. Magnitude of correlation coefficient between predator putative causal variables those predicted by salmon recruitment levels is, on average higher than human or ocean variables. indicating that several predator pcv left a strong imprint on salmon dynamics

Within each domain, how do the various categories stack up

```
# roughly similar mean rho: ssl slightly higher

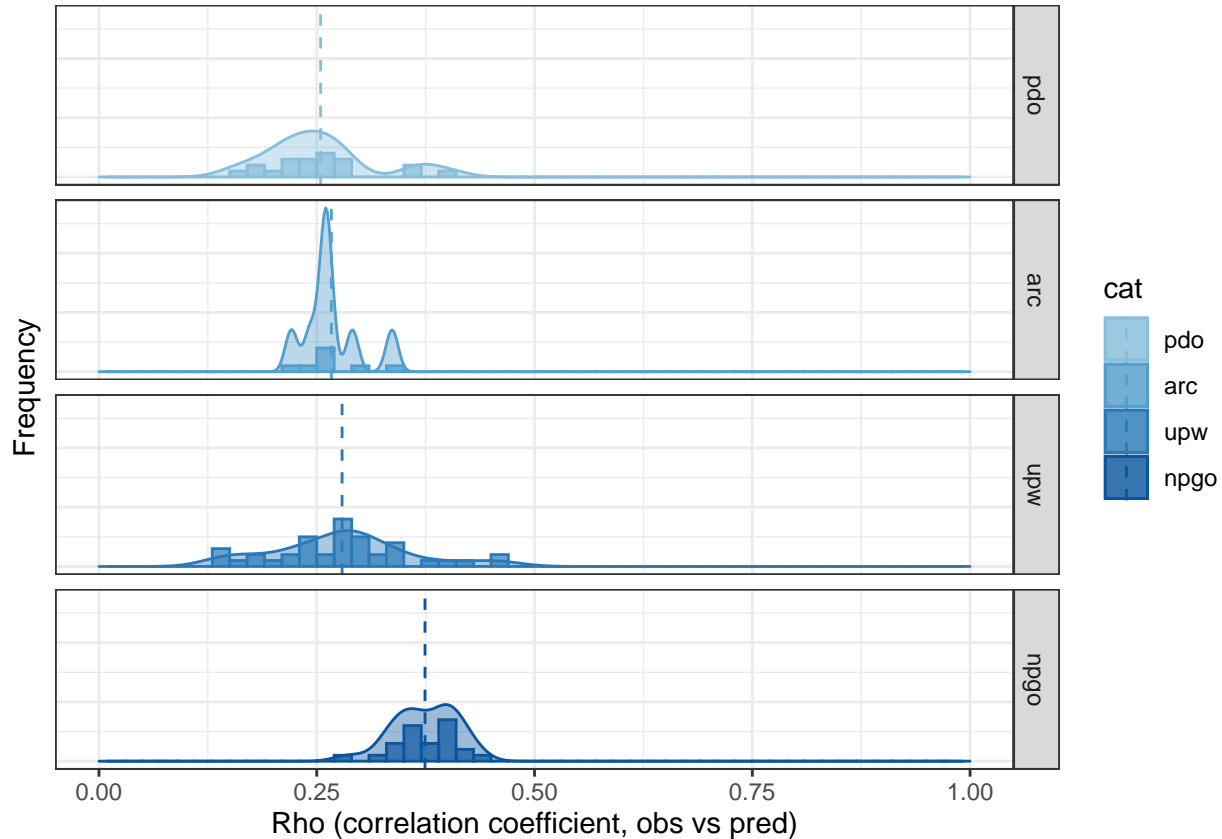
ocean <- df %>%
  filter(domain == "ocean") %>%
  group_by(cat) %>%
  mutate(mu = mean(ESU_rec4n, na.rm=TRUE)) %>%
  ungroup()

ocean$cat = factor(ocean$cat, levels=c('pdo','arc','upw','npgo'))

ggplot(ocean, aes(x = ESU_rec4n, color = cat, fill = cat)) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.7) +
  scale_color_manual(values = colorRampPalette(brewer.pal(9, "Blues"))(8)[4:7]) +
  scale_fill_manual(values = colorRampPalette(brewer.pal(9, "Blues"))(8)[4:7]) +
  geom_vline(aes(xintercept = mu, color = cat),
    linetype = "dashed") +
  geom_density(alpha = 0.4) + facet_grid(cat ~ .) +
  theme_bw() +
  theme(axis.ticks.y = element_blank(),
```

```
axis.text.y = element_blank()) +
ylab("Frequency") + xlab("Rho (correlation coefficient, obs vs pred)") +
xlim(c(0, 1))
```

Warning: Removed 8 rows containing missing values (geom_bar).



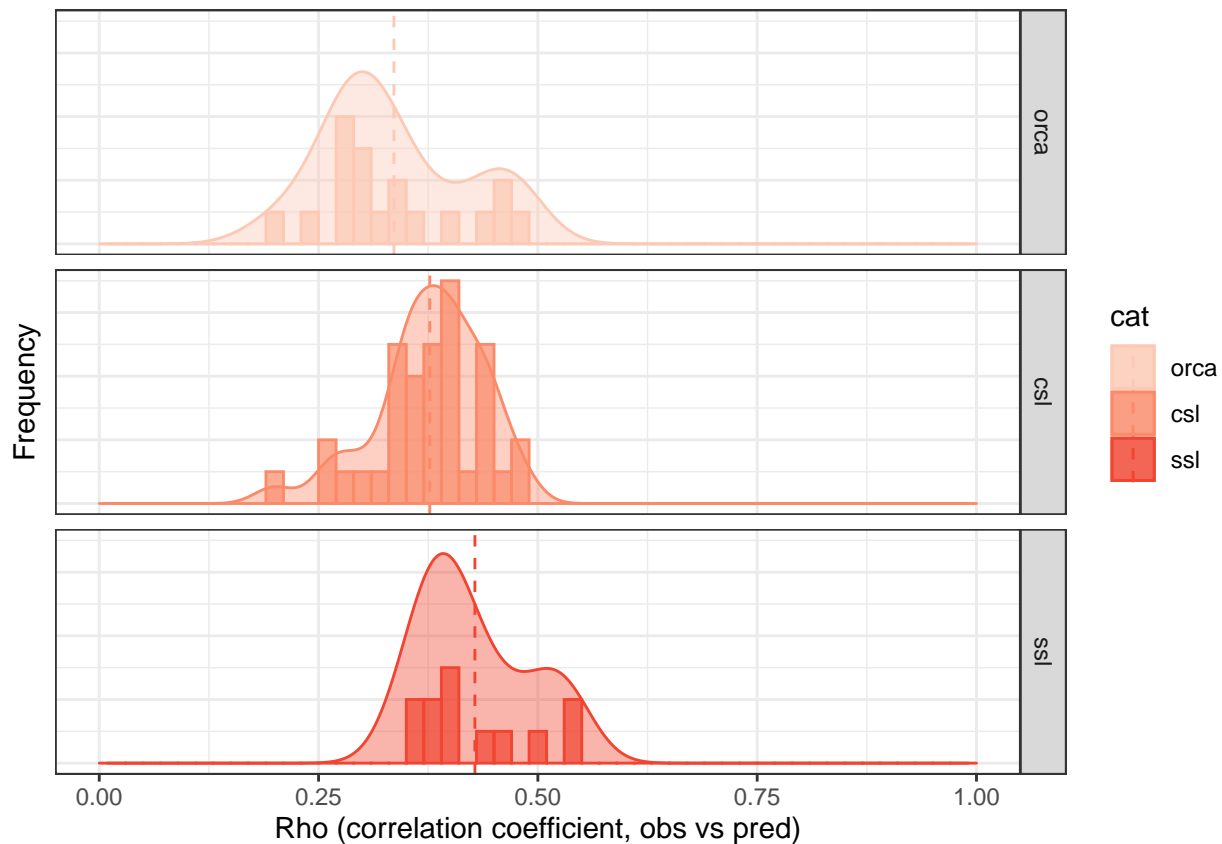
For rec4, NPGO higher than other ocean vars

```
predator <- df %>%
  filter(domain == "predator") %>%
  group_by(cat) %>%
  mutate(mu = mean(ESU_rec4n, na.rm=TRUE)) %>%
  ungroup()

predator$cat = factor(predator$cat, levels=c('orca','csl','ssl'))

ggplot(predator, aes(x = ESU_rec4n, color = cat, fill = cat)) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.7) +
  scale_color_manual(values = colorRampPalette(brewer.pal(9, "Reds"))(6)[2:4]) +
  scale_fill_manual(values = colorRampPalette(brewer.pal(9, "Reds"))(6)[2:4]) +
  geom_vline(aes(xintercept = mu, color = cat),
    linetype = "dashed") +
  geom_density(alpha = 0.4) + facet_grid(cat ~ .) +
  theme_bw() +
  theme(axis.ticks.y = element_blank(),
    axis.text.y = element_blank()) +
  ylab("Frequency") + xlab("Rho (correlation coefficient, obs vs pred)") +
  xlim(c(0, 1))
```

```
## Warning: Removed 8 rows containing non-finite values (stat_bin).
## Warning: Removed 8 rows containing non-finite values (stat_density).
## Warning: Removed 6 rows containing missing values (geom_bar).
```



```
#### For rec4, SSL higher than other pred vars
```

```
human <- df %>%
  filter(domain == "human") %>%
  group_by(cat) %>%
  mutate(mu = mean(ESU_rec4n, na.rm=TRUE)) %>%
  ungroup()

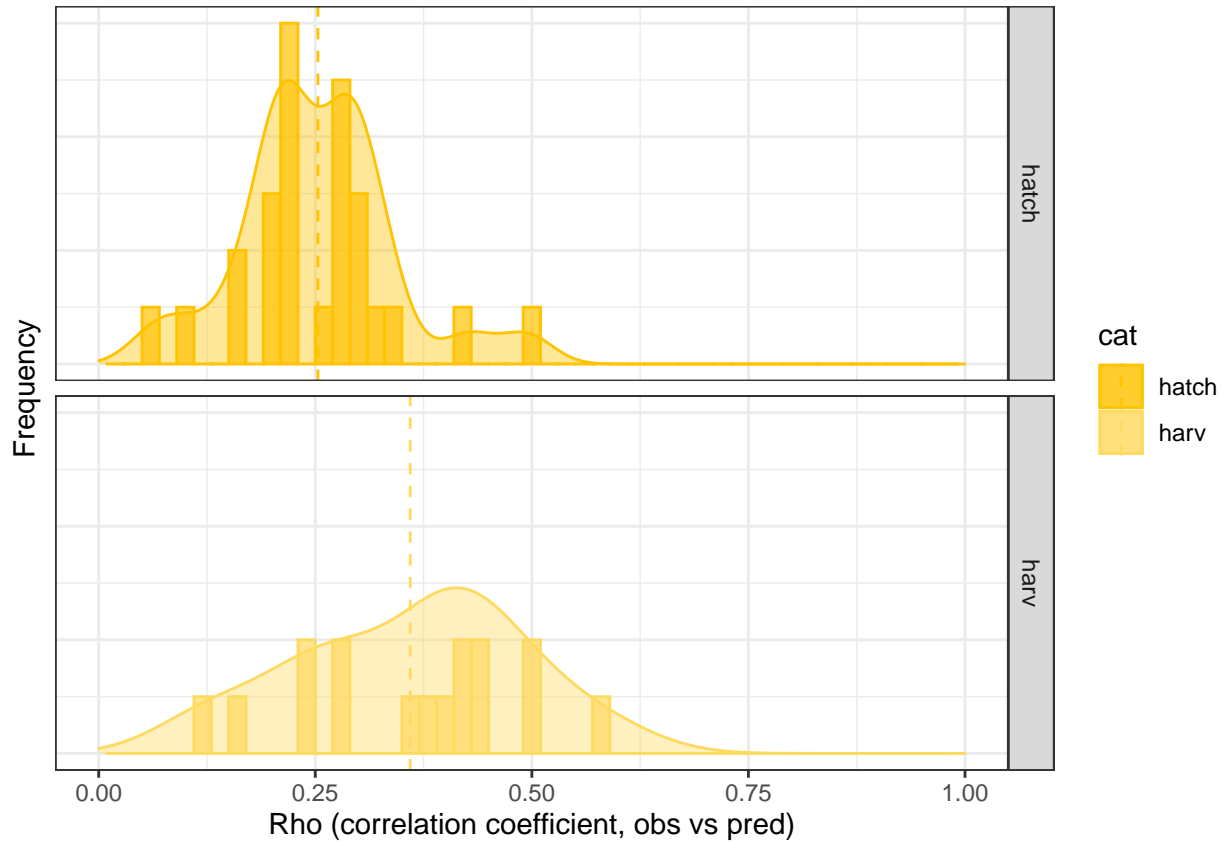
human$cat = factor(human$cat, levels=c('hatch','harv'))

ggplot(human, aes(x = ESU_rec4n, color = cat, fill = cat)) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.7) +
  scale_color_manual(values = c("#FFC300", "#FFDA60")) +
  scale_fill_manual(values = c("#FFC300", "#FFDA60")) +
  geom_vline(aes(xintercept = mu, color = cat),
    linetype = "dashed") +
  geom_density(alpha = 0.4) + facet_grid(cat ~ .) +
  theme_bw() +
  theme(axis.ticks.y = element_blank(),
    axis.text.y = element_blank()) +
  ylab("Frequency") + xlab("Rho (correlation coefficient, obs vs pred)") +
  xlim(c(0, 1))
```

```
## Warning: Removed 14 rows containing non-finite values (stat_bin).
```

Warning: Removed 14 rows containing non-finite values (stat_density).

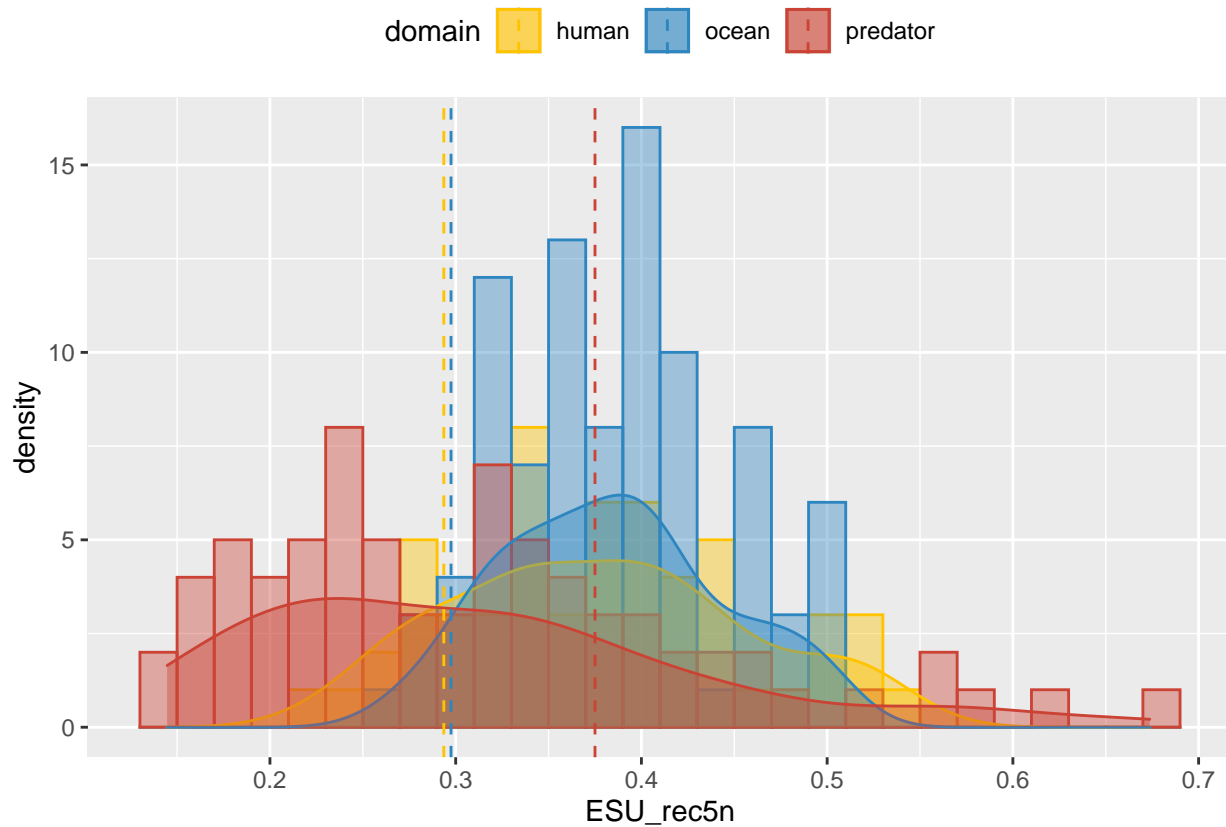
Warning: Removed 4 rows containing missing values (geom_bar).



For rec4, harvest higher than hatchery

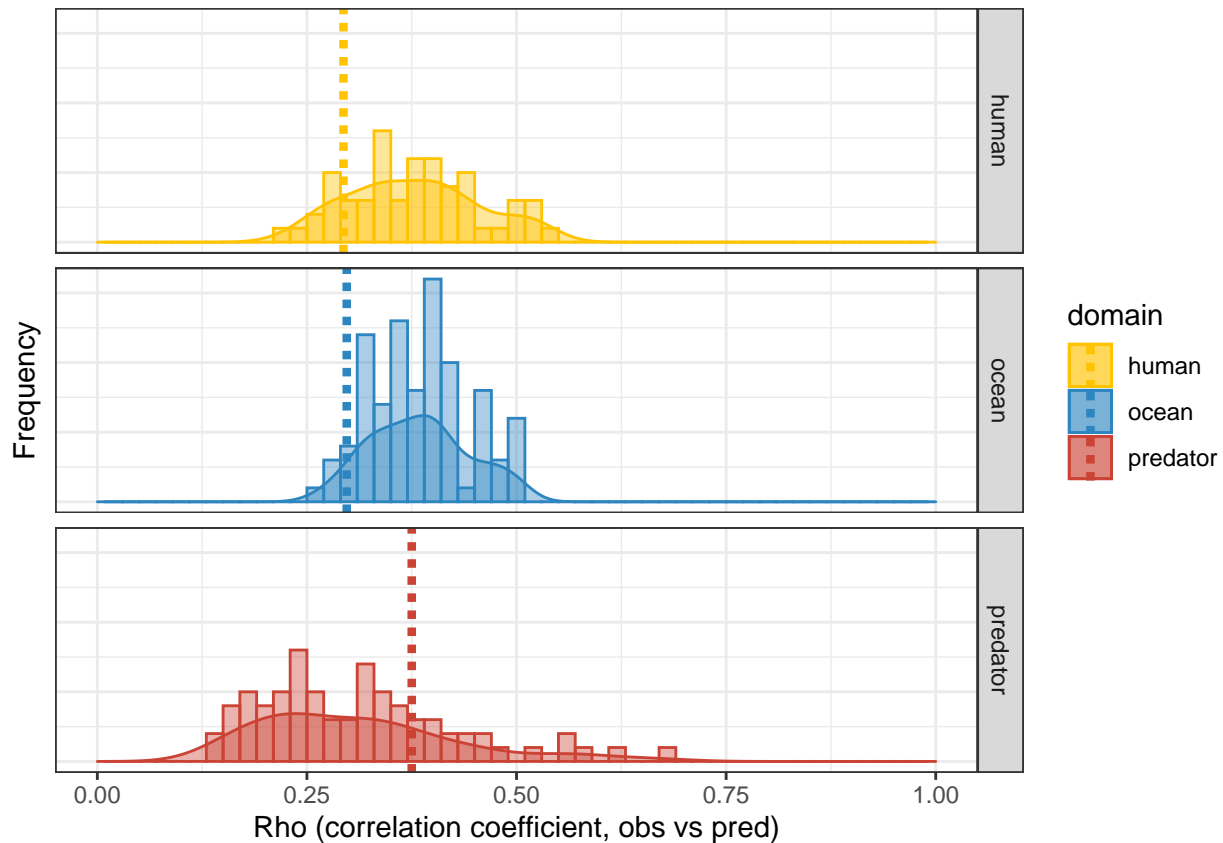
Rec 5

```
ggplot(df, aes(x = ESU_rec5n, color = domain, fill = domain)) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.4) +
  geom_density(alpha = 0.4) +
  geom_vline(aes(xintercept = mu_ESU, color = domain),
             linetype = "dashed") +
  theme(legend.position = "top") +
  scale_color_manual(values=c("#FFC300", "#2E86C1", "#CB4335")) +
  scale_fill_manual(values=c("#FFC300", "#2E86C1", "#CB4335"))
```

```
ggplot(df, aes(x = ESU_rec5n, color = domain, fill = domain )) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.4) +
  geom_density(alpha = 0.4) +
  scale_color_manual(values=c("#FFC300", "#2E86C1", "#CB4335")) +
  scale_fill_manual(values=c("#FFC300", "#2E86C1", "#CB4335")) +
  facet_grid(domain ~ .) +
  geom_vline(aes(xintercept = mu_ESU, color = domain),
             lty = "11", lwd = 1.5) +
  theme_bw() +
  theme(axis.ticks.y = element_blank(),
        axis.text.y = element_blank()) +
  ylab("Frequency") + xlab("Rho (correlation coefficient, obs vs pred)") +
  xlim(c(0, 1))
```

Warning: Removed 6 rows containing missing values (geom_bar).



Compared to rec4, the rec5 mean rho for predators is similar but there is far more right-skew, or a small number of predator vars with very high rh0

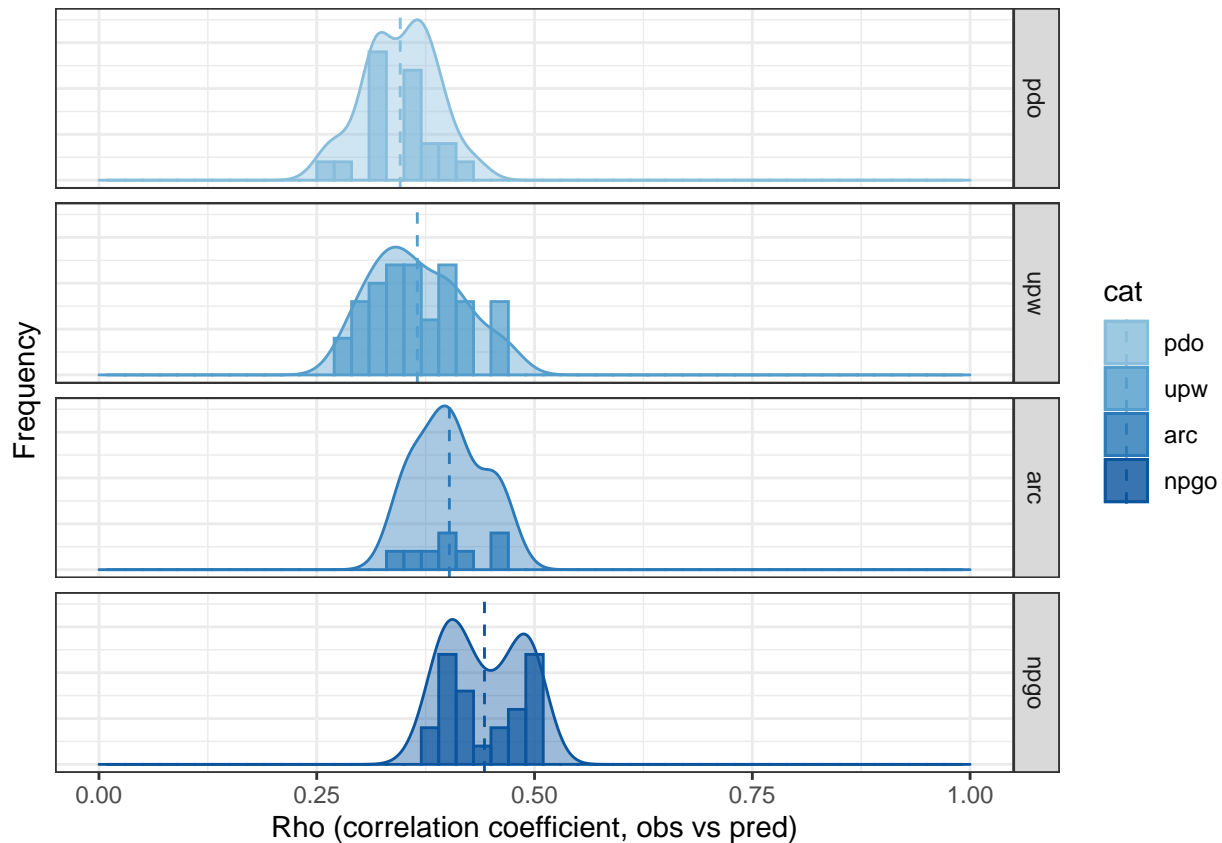
Within each domain, how do the various categories stack up

```
ocean <- df %>%
  filter(domain == "ocean") %>%
  group_by(cat) %>%
  mutate(mu = mean(ESU_rec5n, na.rm=TRUE)) %>%
  ungroup()

ocean$cat = factor(ocean$cat, levels=c('pdo','upw','arc','npgo'))

ggplot(ocean, aes(x = ESU_rec5n, color = cat, fill = cat)) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.7) +
  scale_color_manual(values = colorRampPalette(brewer.pal(9, "Blues"))(8)[4:7]) +
  scale_fill_manual(values = colorRampPalette(brewer.pal(9, "Blues"))(8)[4:7]) +
  geom_vline(aes(xintercept = mu, color = cat),
    linetype = "dashed") +
  geom_density(alpha = 0.4) + facet_grid(cat ~ .) +
  theme_bw() +
  theme(axis.ticks.y = element_blank(),
    axis.text.y = element_blank()) +
  ylab("Frequency") + xlab("Rho (correlation coefficient, obs vs pred)") +
  xlim(c(0, 1))
```

```
## Warning: Removed 8 rows containing missing values (geom_bar).
```



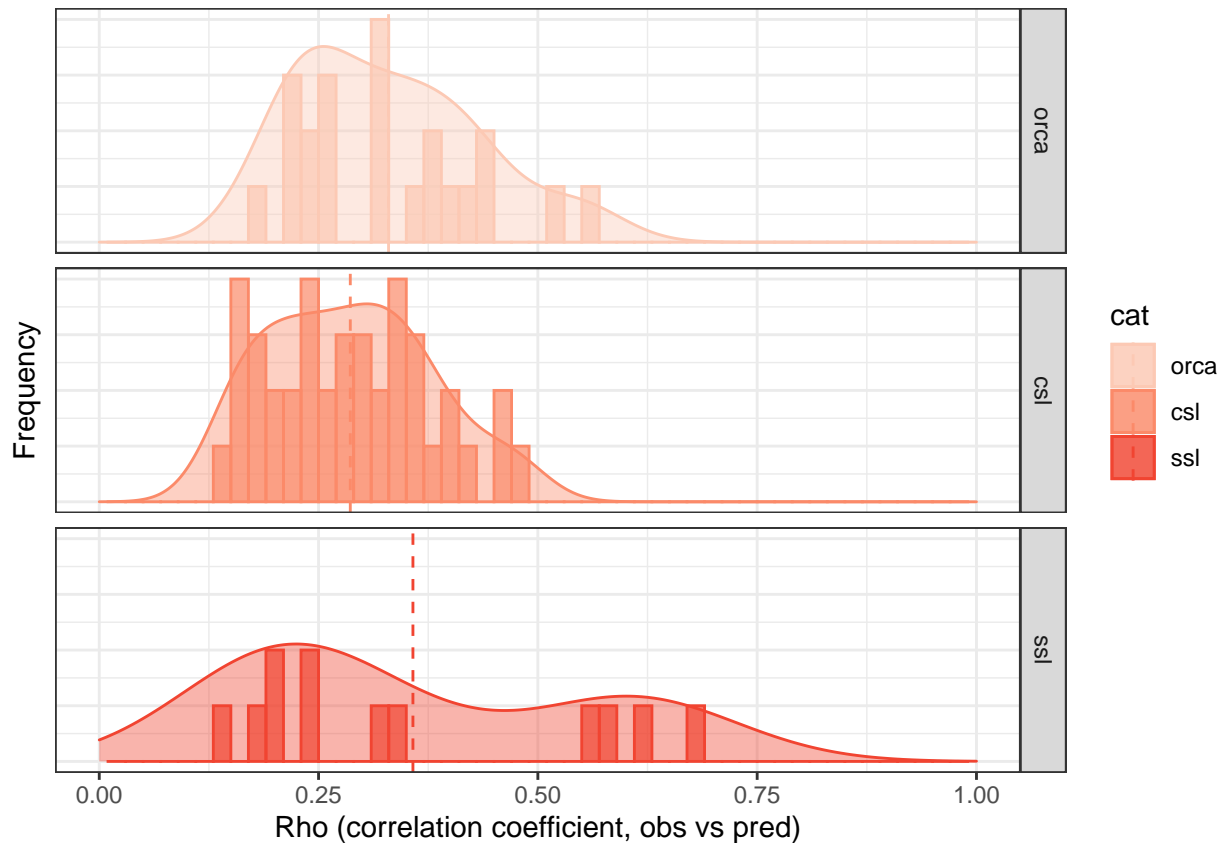
For rec5, NPGO remains higher than other ocean vars

```
predator <- df %>%
  filter(domain == "predator") %>%
  group_by(cat) %>%
  mutate(mu = mean(ESU_rec5n, na.rm=TRUE)) %>%
  ungroup()

predator$cat = factor(predator$cat, levels=c('orca','csl','ssl'))

ggplot(predator, aes(x = ESU_rec5n, color = cat, fill = cat)) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.7) +
  scale_color_manual(values = colorRampPalette(brewer.pal(9, "Reds"))(6)[2:4]) +
  scale_fill_manual(values = colorRampPalette(brewer.pal(9, "Reds"))(6)[2:4]) +
  geom_vline(aes(xintercept = mu, color = cat),
    linetype = "dashed") +
  geom_density(alpha = 0.4) + facet_grid(cat ~ .) +
  theme_bw() +
  theme(axis.ticks.y = element_blank(),
    axis.text.y = element_blank()) +
  ylab("Frequency") + xlab("Rho (correlation coefficient, obs vs pred)") +
  xlim(c(0, 1))
```

Warning: Removed 6 rows containing missing values (geom_bar).



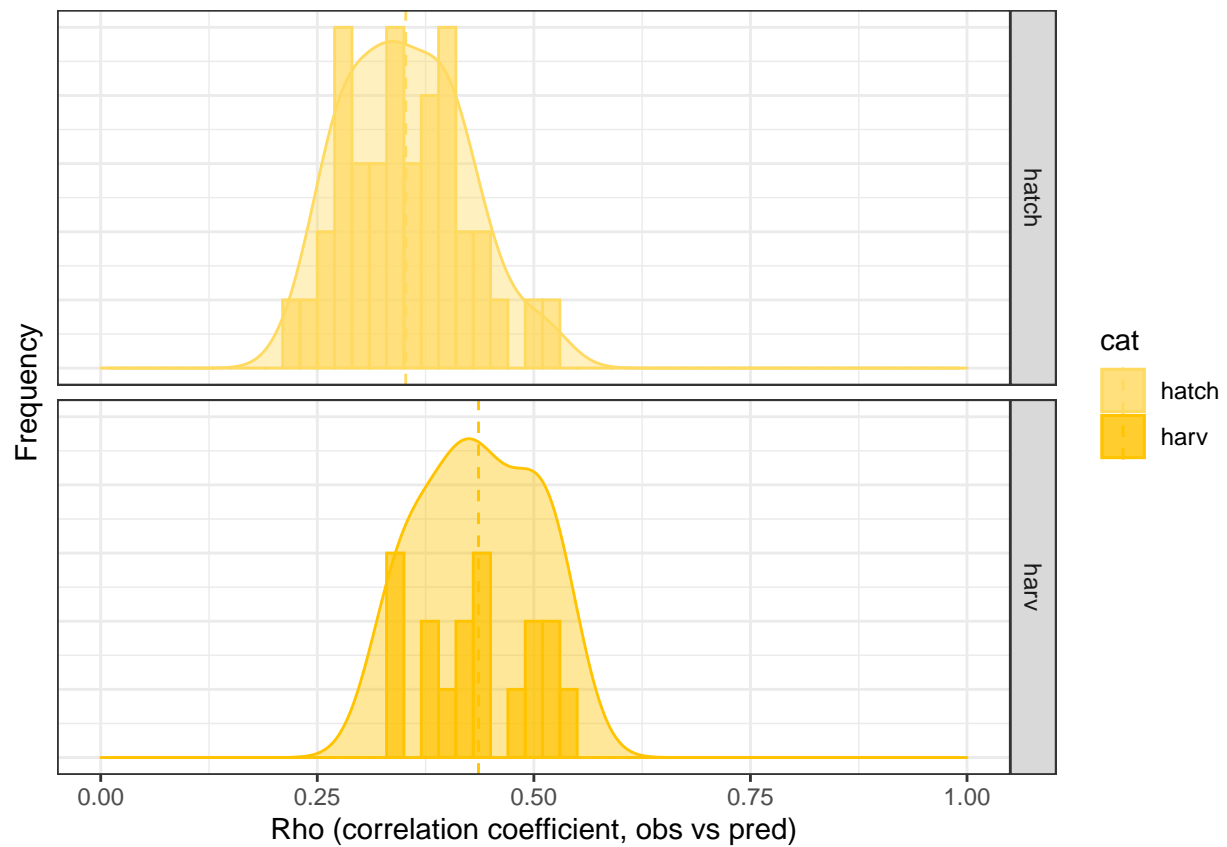
For rec4, SSL remains higher on average than other pred vars, but driven by a small number of variables with $0.5 < \rho < 0.75$ among the highest correlation coefficients observed

```
human <- df %>%
  filter(domain == "human") %>%
  group_by(cat) %>%
  mutate(mu = mean(ESU_rec5n, na.rm=TRUE)) %>%
  ungroup()

human$cat = factor(human$cat, levels=c('hatch','harv'))

ggplot(human, aes(x = ESU_rec5n, color = cat, fill = cat)) +
  geom_histogram(position = "identity", binwidth = 0.02, alpha = 0.7) +
  scale_color_manual(values = c("#FFDA60", "#FFC300")) +
  scale_fill_manual(values = c("#FFDA60", "#FFC300")) +
  geom_vline(aes(xintercept = mu, color = cat),
             linetype = "dashed") +
  geom_density(alpha = 0.4) + facet_grid(cat ~ .) +
  theme_bw() +
  theme(axis.ticks.y = element_blank(),
        axis.text.y = element_blank()) +
  ylab("Frequency") + xlab("Rho (correlation coefficient, obs vs pred)") +
  xlim(c(0, 1))
```

Warning: Removed 4 rows containing missing values (geom_bar).



For rec5, harvest remains higher than hatchery and both show mean magnitudes > those of rec4

Next steps → explore MPGs and compare with ESU