```r
# Workflow to reproduce the cooccurrence analysis in "A new metric for measuring
# the conservancy of plants," Gard et al 2025
# Code was run 21 Jan 2025 for that paper; updated results will differ slightly

library(tidyverse)
library(fqar)

# Download data ------------------------------------------------------------

chic_db <- download_database(80)
chic_inv <- database_inventory(chic_db)
chic_inv_native <- filter(chic_inv, nativity == "native")

chic_assess <- download_assessment_list(80)
chic_trans <- download_transect_list(80)

assess_invs <- assessment_list_inventory(chic_assess)
trans_invs <- transect_list_inventory(chic_trans) # different format

# Align formats of inventories ---------------------------------------------

chic_names <- select(chic_inv,
                     scientific_name,
                     common_name) # for use in trans_adjuster

trans_adjuster <- function(df){
  df %>% rename(scientific_name = species) %>%
    select(scientific_name:duration)  %>%
    left_join(chic_names, by = "scientific_name")
}

trans_invs_adj <- lapply(trans_invs, trans_adjuster) # inventories should now have common
format
all_invs <- c(assess_invs, trans_invs_adj)

# Create cooccurrence database and summary ---------------------------------

chic_co_all <- assessment_cooccurrences(all_invs)
# write_csv(chic_co_all, "chicago_cooccurrences_full.csv") # uncomment to save data
locally

chic_co_sum <- assessment_cooccurrences_summary(all_invs) %>%
  filter(target_species_nativity == "native") # exclude non-native
# 1350 native species vs 1467 in db 149, about 72% vs 75%. Not bad.

sum(chic_co_sum$target_species_n) # 19,530 total occurrences vs 22,259 in 149
sum(chic_co_sum$cospecies_native_n) # 3,185,022 total co-occurrences vs 1,207,772 in 149
(wow!)

chic_sm <- chic_co_sum %>%
  filter(target_species_n >= 3) # down to 918

# write_csv(chic_sm, "chicago_cooccurrences_summary.csv") # uncomment to save data locally

# Analysis -----------------------------------------------------------------

ggplot(chic_sm, aes(y = as.factor(target_species_c),
                    x = cospecies_native_mean_c)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(alpha = .18) +
  theme_minimal() +
  labs(x = "CNMC", y = "Assigned C-value")

cor.test(chic_sm$target_species_c, chic_sm$cospecies_native_mean_c)
```

```r
model <- lm(target_species_c ~ cospecies_native_mean_c,
            data = chic_sm)
summary(model)

range(chic_sm$cospecies_native_mean_c)

# estimate target_C = -11.45 + 3.26 * CNMC but note limits on range

ggplot(chic_sm, aes(y = target_species_c,
                    x = cospecies_native_mean_c)) +
  geom_jitter(alpha = .3) +
  geom_smooth(method = "lm") +
  theme_minimal()

# some data sets included in paper ----------------------------------------

chic_smaller <- chic_sm %>%
  select(target_species,
         target_species_c,
         target_species_n,
         cospecies_native_mean_c,
         cospecies_native_n,
         discrepancy_c)

overrated <- chic_smaller %>%
  slice_max(order_by = discrepancy_c,
            n = 10) %>%
  arrange(-discrepancy_c)

underrated <- chic_smaller %>%
  slice_min(order_by = discrepancy_c,
            n = 10) %>%
  arrange(discrepancy_c)

# profile plots -----------------------------------------------------------

species_profile_plot("Epilobium ciliatum",
                     all_invs,
                     native = TRUE)
species_profile_plot("Solidago canadensis",
                     all_invs,
                     native = TRUE)
```