

# tidymodels-m333

Grayson White

```
# Import data
spatial <- read_csv("../data/plot_level/plt_spatial.csv")
response <- read_csv("../data/plot_level/plot_response.csv")

# Join data
## Keep only observations in both 'spatial' and 'response'
dat <- inner_join(spatial, response,
                  by = c("PLT_CN" = "PLT_CN",
                        "INVYR" = "INVYR"))

# Create columns for province, sections, and subsections
dat <- dat %>%
  mutate(
    subsection = ECOSUBCD.x,
    section = str_remove_all(ECOSUBCD.x, "[:lower:]"),
    province = str_sub(section, end = -2)
  )

# Select small subset of columns to work with for this EDA
dat_small <- dat %>%
  dplyr::select(PLT_CN, INVYR, PLOT.x, LON_PUBLIC.x, LAT_PUBLIC.x, LON_PUBLIC.y, LAT_PUBLIC.y,
                ELEV_PUBLIC.x, ELEV_PUBLIC.y, forgrp, forprob, nlcd11, demLF, evtLF, forbio,
                BALIVE_TPA, CNTLIVE_TPA, BIOLIVE_TPA, VOLNLIVE_TPA, subsection, section, province)

# Remove redundant columns, rename columns for ease of use
dat_small <- dat_small %>%
  dplyr::select(-LON_PUBLIC.y, -LAT_PUBLIC.y, -ELEV_PUBLIC.y) %>%
  rename(PLOT = PLOT.x,
         LON_PUBLIC = LON_PUBLIC.x,
         LAT_PUBLIC = LAT_PUBLIC.x,
         ELEV_PUBLIC = ELEV_PUBLIC.x)

north_rocky <- dat_small %>%
  filter(province == "M333")
```

## Fit a model

```
set.seed(1)
hier_model_spec <- linear_reg() %>%
  set_engine("stan-glmer")

nr_train <- north_rocky %>%
```

```

sample_frac(0.75)

nr_test <- north_rocky %>%
  anti_join(nr_train)

## Joining, by = c("PLT_CN", "INVYR", "PLOT", "LON_PUBLIC", "LAT_PUBLIC", "ELEV_PUBLIC", "forgrp", "for")

hier_model_fit <-
  hier_model_spec %>%
  fit(BIOLIVE_TPA ~ 1 + forprob + (1 | subsection),
      data = nr_train)

hier_model_fit

## parsnip model object
##
## Fit time: 39.8s
## stan_glmer
## family: gaussian [identity]
## formula: BIOLIVE_TPA ~ 1 + forprob + (1 | subsection)
## observations: 2252
## -----
##               Median MAD_SD
## (Intercept) -6.0      4.1
## forprob      50.4      3.5
##
## Auxiliary parameter(s):
##               Median MAD_SD
## sigma 35.4      0.6
##
## Error terms:
## Groups      Name      Std.Dev.
## subsection (Intercept) 12
## Residual              35
## Num. levels: subsection 20
##
## -----
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg

results_test <- hier_model_fit %>%
  predict(new_data = nr_test) %>%
  mutate(
    truth = nr_test$BIOLIVE_TPA
  )

results_test %>%
  rmse(truth = truth, estimate = .pred)

## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>       <dbl>
## 1 rmse   standard      35.3

```

```

results_train <- hier_model_fit %>%
  predict(new_data = nr_train) %>%
  mutate(
    truth = nr_train$BIOLIVE_TPA
  )

results_train %>%
  rmse(truth = truth, estimate = .pred)

```

```

## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>       <dbl>
## 1 rmse    standard      35.3

```

```

nr_fit <-
  hier_model_spec %>%
  fit(BIOLIVE_TPA ~ 1 + forprob + (1 | subsection),
      data = north_rocky)

model_df <- data.frame(fitted = nr_fit$fit$fitted.values,
                      true = north_rocky$BALIVE_TPA,
                      subsection = north_rocky$subsection)

model_df %>%
  group_by(subsection) %>%
  summarize(mean_fit = mean(fitted),
            mean_true = mean(true)) %>%
  ggplot(aes(x = subsection,
            y = mean_fit)) +
  geom_point() +
  geom_point(aes(y = mean_true), color = "red") +
  theme_bw() +
  theme(axis.text.x = element_text(
    angle = 90,
    vjust = 0.5,
    hjust = 1
  )) +
  labs(
    x = "Subsection",
    y = "Estimated (Black) and True (Red) Average Biomass"
  )

```

```

## 'summarise()' ungrouping output (override with '.groups' argument)

```

