

# Activity 4: Status report on semester project

Karthik Thrikkadeeri

2025-04-10

## 1 Background

2 Ecologists have long been fascinated by patterns of biodiversity along environmental gradients  
3 and the mechanisms driving these patterns (Ricklefs 2004). For the specific question of how  
4 communities of organisms are assembled with certain species, there are two main factors gener-  
5 ally believed to play a role: environmental filtering and biotic interactions (Cavender-Bares et  
6 al. 2009). Considering the functional and phylogenetic structure of communities can provide  
7 insight into the relative importance of these drivers of community assembly (Webb et al. 2002;  
8 Cavender-Bares et al. 2009). Particularly, assemblages of species more clustered or dispersed  
9 in function or phylogeny than expected by chance can hint at strong effects of environmental  
10 filtering or competitive interactions (Lebrija-Trejos et al. 2010; Mayfield and Levine 2010).  
11 Elevational gradients are interesting from this perspective, due to their rapid turnover of envi-  
12 ronmental conditions creating complex sets of habitats and climate zones (Graham et al. 2014).  
13 Previous studies on this topic have focused on taxonomic diversity of community assemblages  
14 and found that species richness either decreases with elevation or peaks at mid-elevations (e.g.,

15 McCain 2009), while the limited number of studies focusing on functional and phylogenetic  
16 diversity have found decreasing trends within single mountains (e.g., Hanz et al. 2019). To  
17 test for the generality of these patterns and their deterministic strengths, Montaña-Centellas  
18 et al. (2020) analysed a large dataset of resident birds of 46 mountains across the globe.

## 19 **Process**

20 The process of obtaining the data for this (re)analysis was challenging because the authors  
21 have not provided the raw datasets, but also simple at the same time because the processed  
22 dataset has been made available on [Dryad](#). I simply downloaded the single data file and  
23 stored it in my repository at `data/montano_dryad_data.xlsx`. This file contains functional  
24 and phylogenetic richness and diversity metrics calculated at the level of 100 m elevation bands  
25 across 46 mountains of the world, along with their SESs and corresponding p-values.

26 Of the three main results from the paper I wanted to reproduce, I will not be able to reproduce  
27 the final analysis of the effect of latitude on patterns of elevational change, because latitude  
28 information is missing from the dataset available. The other two seem feasible to reproduce,  
29 but I have so far only been able to reproduce the first, quadratic regressions for patterns of bird  
30 functional and phylogenetic diversity along elevational gradients of the world. Although the  
31 authors mention “classifying” each elevational pattern using slope and intercept values from  
32 the model results, they do not provide code for this which suggests that they used manual  
33 classification. I have tried to achieve the same results using a programmatic approach. For

34 the second result, again, since we do not have access to the original trait and phylogeny data,  
35 I am limited to using SES values that they have already calculated for each metric. This  
36 means many of the intermediate steps are ambiguous, and I can only assume reproducibility  
37 and absence of errors.

38 One R package in this analysis was new to me, the `metafor` package (Viechtbauer 2010) used  
39 to run hierarchical multivariate meta-regression testing for the effect of latitude on patterns  
40 of deterministic signal. Aside from this, the authors also used `lme4` (Bates et al. 2015) and  
41 `stats` (R Core Team 2025) packages for the analyses. Whereas, for reproducing their analysis,  
42 I am currently using the following additional packages: `dplyr` (Wickham et al. 2023), `readxl`  
43 (Wickham and Bryan 2023), `glue` (Hester and Bryan 2024), `ggplot2` (Wickham 2016), `purrr`  
44 (Wickham and Henry 2023), and `broom` (Robinson et al. 2024).

## 45 **Reproducible research code**

46 The below code sets up everything required to run the analyses:

```
library(dplyr)

library(readxl) # file is .xlsx

library(glue)

library(ggplot2); theme_set(theme_classic())

library(purrr)

library(broom) # to easily access p-value from model objects
```

```

# colour palette to match that used in paper

scale_col_pal <- scale_color_brewer(palette = "Dark2", na.value = "grey50")
scale_fill_pal <- scale_fill_brewer(palette = "Dark2", na.value = "grey50")

data_dryad <- read_xlsx("data/montano_dryad_data.xlsx",
                        sheet = "FDPD_Supporting")

# load functions for classification

source("sem-project/class_functions.R")

```

47 For the first analysis to obtain patterns of diversity metrics across elevations, I first need to fit  
 48 the polynomial model for each of the four diversity metrics across all 46 mountains individually,  
 49 and then extract the intercept and coefficients for each as well as predict responses across  
 50 elevations.

```

# iterate over each mountain in the dataset:

# fit LM of each metric against linear and second-order polynomial terms of Elevation

# extract intercept and coefficients for both Elevation terms

# aggregate all these values over all mountains

list_mount <- unique(data_dryad$Mountain)

metric_cols <- c("FRic", "FDis", "PD", "MPD")

```

```

to_iter <- expand.grid(mountain = list_mount, metric = metric_cols)

quad_reg <- map2(to_iter$mountain, to_iter$metric, ~ {

  data_mount <- data_dryad %>% filter(Mountain == .x)

  model_formula <- as.formula(glue("{.y} ~ elevation + I(elevation^2)"))

  model_mount <- lm(model_formula, data_mount)

  # predict only for elevations present in each mountain
  to_pred <- data_mount %>% distinct(Mountain, elevation)

  # predict
  model_pred <- predict(model_mount, newdata = to_pred, type = "response")

  pred <- bind_cols(to_pred, pred = model_pred)

  # store p-value of model fit
  p <- broom::glance(model_mount) %>% pull(p.value)

  t(model_mount$coefficients) %>%

    as_tibble() %>%

    magrittr::set_colnames(c("intercept", "coef_linear", "coef_poly")) %>%

    bind_cols(tibble(mountain = .x, metric = .y, pval = p)) %>%

```

```

    left_join(pred, by = c("mountain" = "Mountain"))

}) %>%

list_rbind() %>%

relocate(metric, mountain) # bring metric and mountain cols to the start

```

51 I then need to plot the predicted values and classify the trends into the nine categories the  
 52 authors used: Increasing, Decreasing, Mid Peak, Mid Valley, Low Plateau (Peak), High Plateau  
 53 (Peak), Low Valley, High Valley, and NS for non-significant.

```

quad_reg %>%

ggplot(aes(x = elevation, y = pred, col = mountain)) +

geom_line() +

facet_wrap(~ metric, ncol = 1, scales = "free_y", strip.position = "left") +

labs(x = "Elevation", y = "") +

# scale_col_pal + # only makes sense to colour scheme after classifying into 8 categs

theme(panel.grid.major.x = element_line(colour = "grey", linetype = "dotted"),

      # match strip format in paper

      strip.background = element_rect(colour = NA),

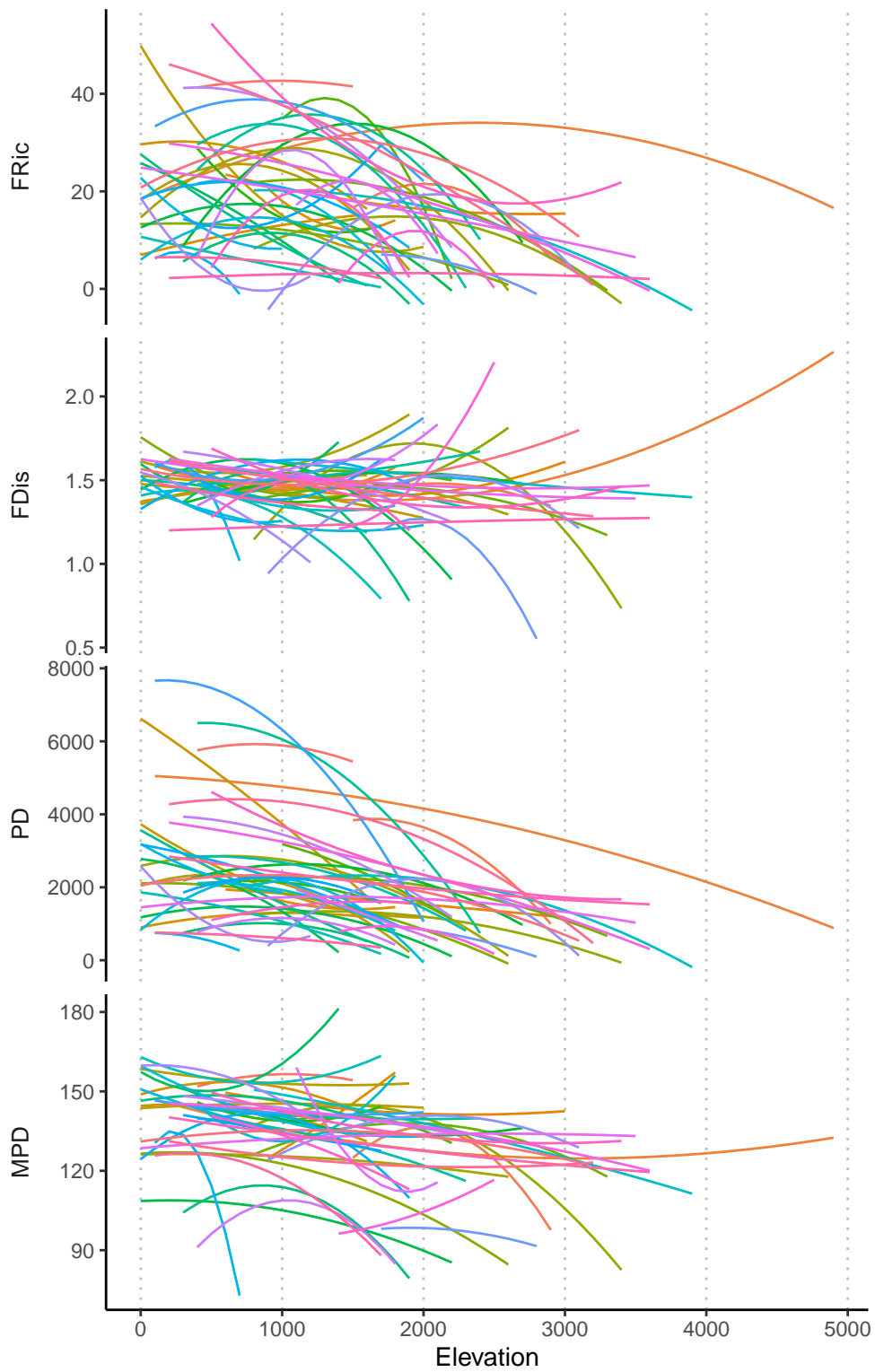
      strip.placement = "outside",

      strip.text = element_text(size = 10),

      # remove colour key

```

```
legend.position = "none")
```





55 The predicted values match up well with what the authors obtained. They do not specify how  
56 exactly they achieved the classification, and it seems to have been done manually. However, I  
57 want to do this programmatically, for which I first create a set of custom functions:

```
# get inflection point

get_inflection <- function(data) {

  data %>%

    group_by(metric, mountain) %>%

    # calculate simple diff

    mutate(pred_min1 = lag(pred, default = first(pred)),

           pred_plus1 = lead(pred, default = last(pred))) %>%

    mutate(pred_diff = pred - pred_min1) %>%

    # decide whether or not that elevational band is an inflection point for the trend

    # only need lead, since we want just one elevation to be inflection point

    # get two TRUEs if use lead and lag

    mutate(pred_diff_plus1 = lead(pred_diff, default = last(pred_diff))) %>%

    mutate(inflection = case_when(

      # if first or last elevational band, not inflection

      (pred_min1 == pred | pred_plus1 == pred) ~ FALSE,
```

```

    (pred_diff_plus1 >= 0 & pred_diff >= 0) |

    (pred_diff_plus1 <= 0 & pred_diff <= 0) ~ FALSE,

    TRUE ~ TRUE

  )) %>%

  # select(-contains("1")) %>%

  select(metric, mountain, elevation, pval, pred, pred_diff, inflection) %>%

  ungroup()

}

# classify elevational bands into L, M, H

get_elev_cats <- function(data) {

  data_cats <- data %>%

    distinct(mountain, elevation) %>%

    group_by(mountain) %>%

    mutate(elev_min = min(elevation),

           elev_max = max(elevation)) %>%

```

```

# could do ceiling and floor for first and second if we want to be conservative, or oppos

# but leave it like this for now

mutate(elev_thresh1 = round(quantile(elevation, probs = 1/3), -2),
       elev_thresh2 = round(quantile(elevation, probs = 2/3), -2)) %>%

group_by(mountain, elevation) %>%

reframe(elev_cat = case_when(
  elevation < elev_thresh1 ~ "LOW",
  elevation > elev_thresh2 ~ "HIGH",
  TRUE ~ "MID"
))

data %>% left_join(data_cats, by = c("mountain", "elevation"))

# Mikura-Jima is the mountain with fewest elevational bands classified as MID

# but n = 3 (> 2) so this is fine

}

# classify trends as valley or peak

```

```

class_valley_peak <- function(data) {

  data_class <- data %>%

    group_by(metric, mountain) %>%

    filter(elev_cat == "LOW") %>%

    reframe(valleypeak = case_when(

      nth(pred_diff, 2) < 0 ~ "valley",

      nth(pred_diff, 2) > 0 ~ "peak",

      TRUE ~ NA

    ))

  data %>% left_join(data_class, by = c("metric", "mountain"))

}

```

58 I then code the classification pipeline:

```

# classify elevational trends

quad_reg_class <- quad_reg %>%

  get_inflection() %>%

  get_elev_cats() %>%

```

```

class_valley_peak() %>%

group_by(metric, mountain) %>%

reframe(trend = case_when(

  min(pval) > 0.1 ~ "NS",

  all(inflection == FALSE) & all(pred_diff >= 0) ~ "Increasing",
  all(inflection == FALSE) & all(pred_diff <= 0) ~ "Decreasing",

  all(valleypeak == "valley") & all(!(inflection == TRUE & elev_cat != "LOW")) ~ "Low Valley",
  all(valleypeak == "valley") & all(!(inflection == TRUE & elev_cat != "HIGH")) ~ "High Valley",
  all(valleypeak == "valley") & all(!(inflection == TRUE & elev_cat != "MID")) ~ "Mid Valley",

  all(valleypeak == "peak") & all(!(inflection == TRUE & elev_cat != "LOW")) ~ "Low Peak",
  all(valleypeak == "peak") & all(!(inflection == TRUE & elev_cat != "HIGH")) ~ "High Peak",
  all(valleypeak == "peak") & all(!(inflection == TRUE & elev_cat != "MID")) ~ "Mid Peak",

  TRUE ~ NA

))

# plot frequency bar graph of different trend types

quad_reg_class %>%

```

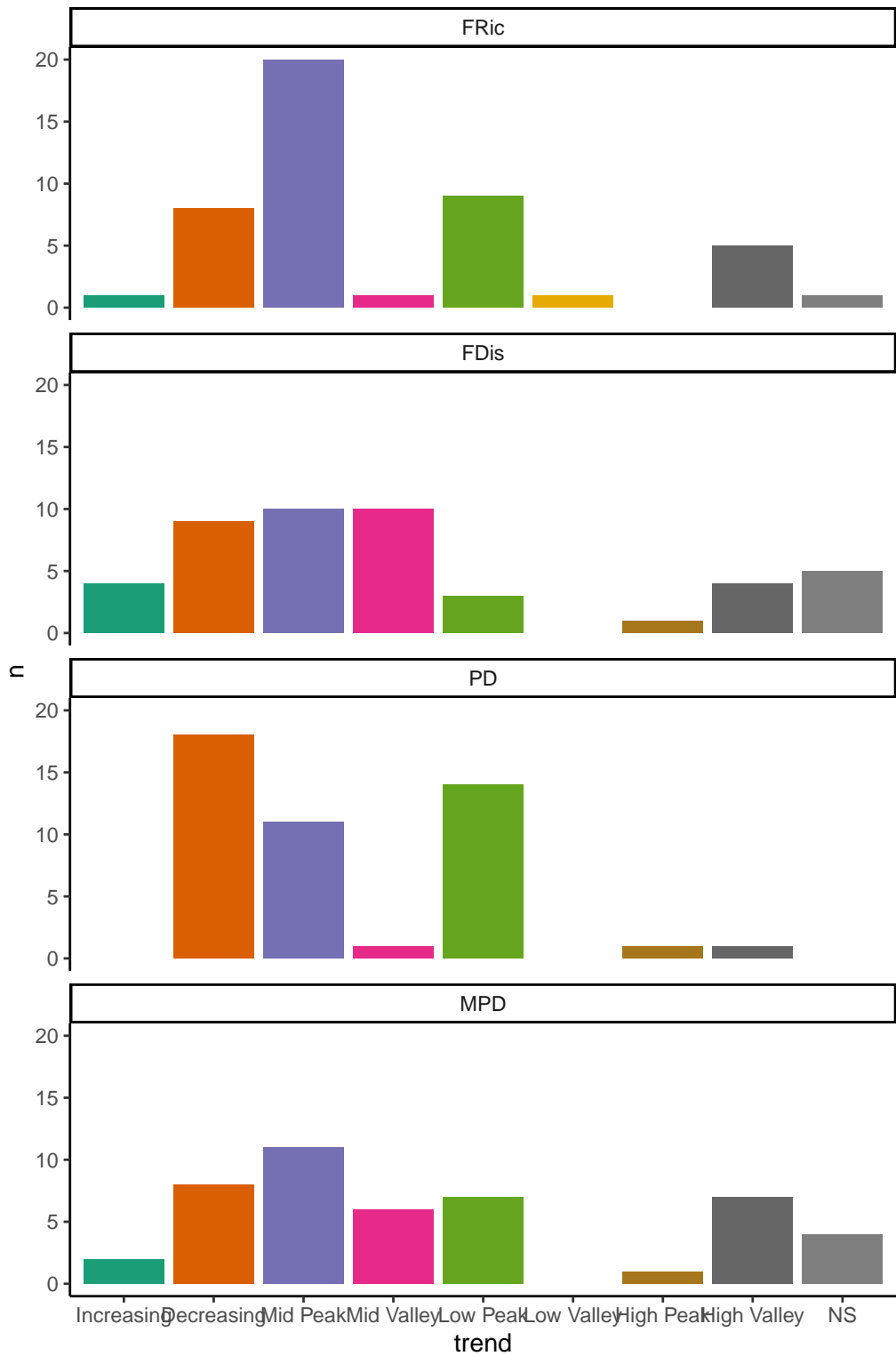
```

mutate(trend = factor(trend,
                      levels = c("Increasing", "Decreasing", "Mid Peak", "Mid Valley",
                                "Low Peak", "Low Valley", "High Peak", "High Valley", "N

group_by(metric) %>%
count(trend) %>%

ggplot() +
geom_col(aes(x = trend, y = n, fill = trend)) +
facet_wrap(~ metric, ncol = 1) +
scale_fill_pal +
theme(legend.position = "none")

```



60 As expected, there is a discrepancy in the frequency bar graph of classified trends, despite the  
61 overall similarity in proportions.

## 62 **Remaining steps and hurdles**

63 Firstly, I need to recheck my programmatic classification method to ensure there are no errors  
64 and that the classification I obtained is sensible. I will also try to understand what exact  
65 differences in methodology have led to this discrepancy. But regardless of this, the broader  
66 inferences might still remain the same.

67 Then, I will work on the second main analysis of the paper, looking at trends in deterministic  
68 strength of these elevational patterns. I do not have any major concerns for this part, since I  
69 do not foresee any steps in the methodology that have not been explicitly mentioned.

## 70 **References**

- 71 Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. [Fitting linear mixed-effects models](#)  
72 [using lme4](#). Journal of Statistical Software 67:1–48.
- 73 Cavender-Bares, J., K. H. Kozak, P. V. Fine, and S. W. Kembel. 2009. The merging of  
74 community ecology and phylogenetic biology. Ecology letters 12:693–715.
- 75 Graham, C. H., A. C. Carnaval, C. D. Cadena, K. R. Zamudio, T. E. Roberts, J. L. Parra,  
76 C. M. McCain, et al. 2014. The origin and maintenance of montane diversity: Integrating  
77 evolutionary and ecological processes. Ecography 37:711–719.



78 Hanz, D. M., K. Böhning-Gaese, S. W. Ferger, S. A. Fritz, E. L. Neuschulz, M. Quitián, V.  
 79 Santillán, et al. 2019. Functional and phylogenetic diversity of bird assemblages are filtered  
 80 by different biotic factors on tropical mountains. *Journal of Biogeography* 46:291–303.  
 81 Hester, J., and J. Bryan. 2024. [Glue: Interpreted string literals](#).  
 82 Lebrija-Trejos, E., E. A. Pérez-García, J. A. Meave, F. Bongers, and L. Poorter. 2010. Func-  
 83 tional traits and environmental filtering drive community assembly in a species-rich tropical  
 84 system. *Ecology* 91:386–398.  
 85 Mayfield, M. M., and J. M. Levine. 2010. Opposing effects of competitive exclusion on the  
 86 phylogenetic structure of communities. *Ecology letters* 13:1085–1093.  
 87 McCain, C. M. 2009. Global analysis of bird elevational diversity. *Global Ecology and Bio-*  
 88 *geography* 18:346–360.  
 89 Montaña-Centellas, F. A., C. McCain, and B. A. Loiselle. 2020. Using functional and phyloge-  
 90 netic diversity to infer avian community assembly along elevational gradients. *Global Ecology*  
 91 *and Biogeography* 29:232–245.  
 92 R Core Team. 2025. [R: A language and environment for statistical computing](#). R Foundation  
 93 for Statistical Computing, Vienna, Austria.  
 94 Ricklefs, R. E. 2004. A comprehensive framework for global patterns in biodiversity. *Ecology*  
 95 *letters* 7:1–15.  
 96 Robinson, D., A. Hayes, and S. Couch. 2024. [Broom: Convert statistical objects into tidy](#)  
 97 [tibbles](#).  
 98 Viechtbauer, W. 2010. [Conducting meta-analyses in R with the metafor package](#). *Journal of*  
 99 *Statistical Software* 36:1–48.

100 Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and  
101 community ecology. *Annual review of ecology and systematics* 33:475–505.

102 Wickham, H. 2016. [ggplot2: Elegant graphics for data analysis](#). Springer-Verlag New York.

103 Wickham, H., and J. Bryan. 2023. [Readxl: Read excel files](#).

104 Wickham, H., R. François, L. Henry, K. Müller, and D. Vaughan. 2023. [Dplyr: A grammar](#)  
105 [of data manipulation](#).

106 Wickham, H., and L. Henry. 2023. [Purrr: Functional programming tools](#).