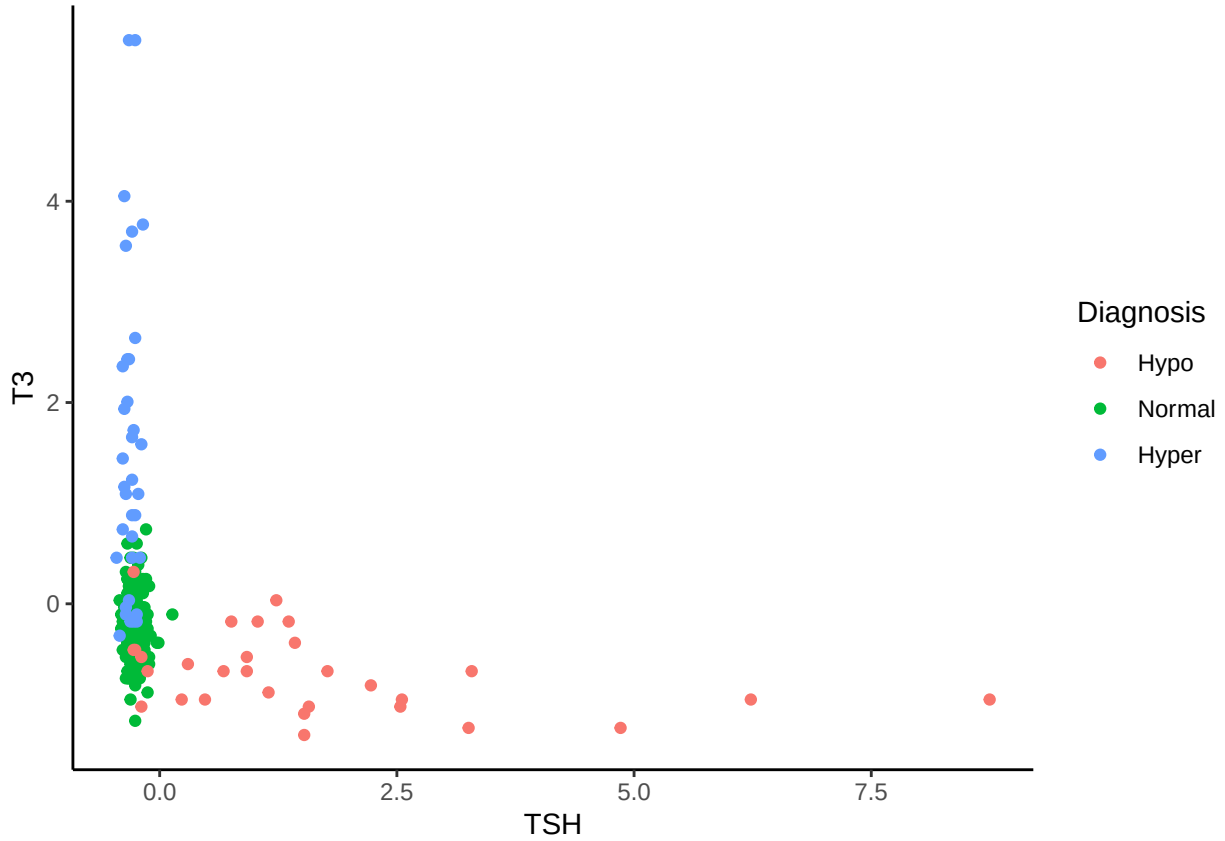
Thyroid Example

Jonathan Klus

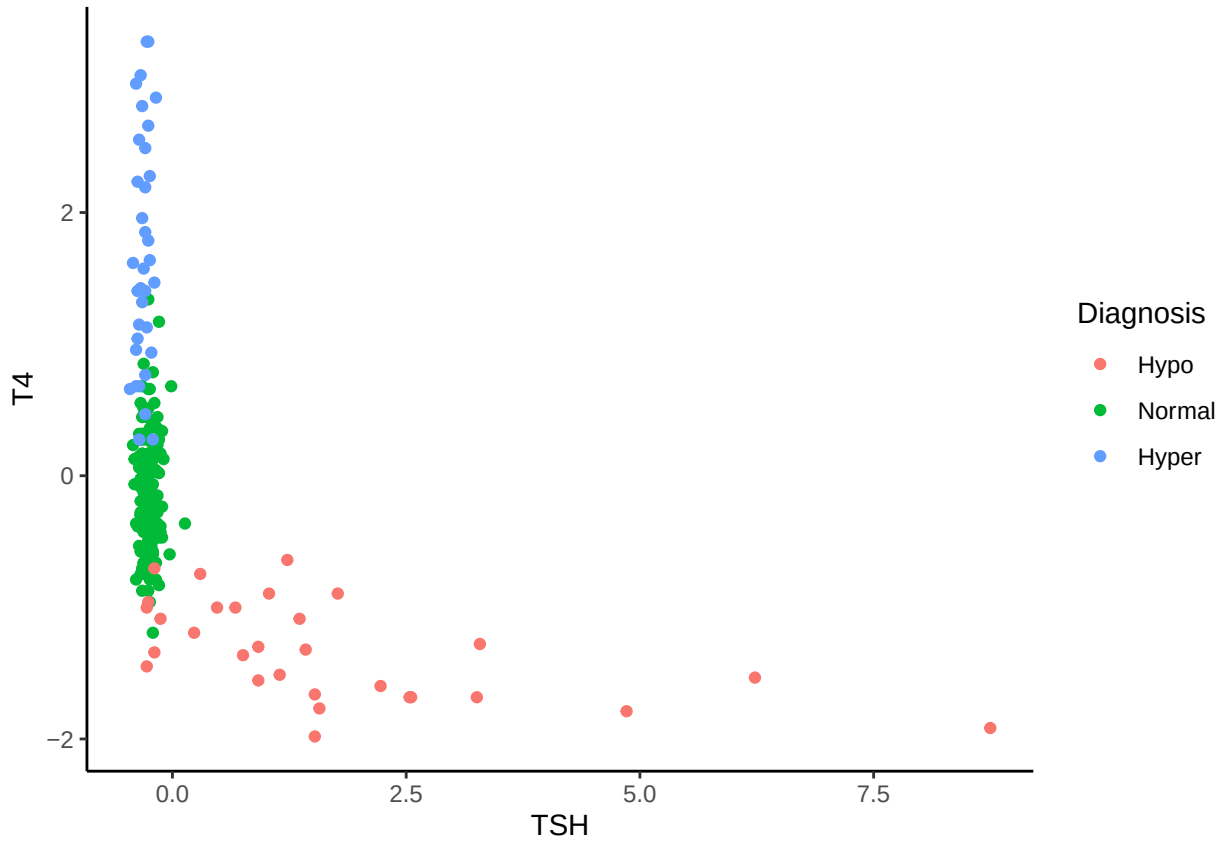
2024-03-01

Data

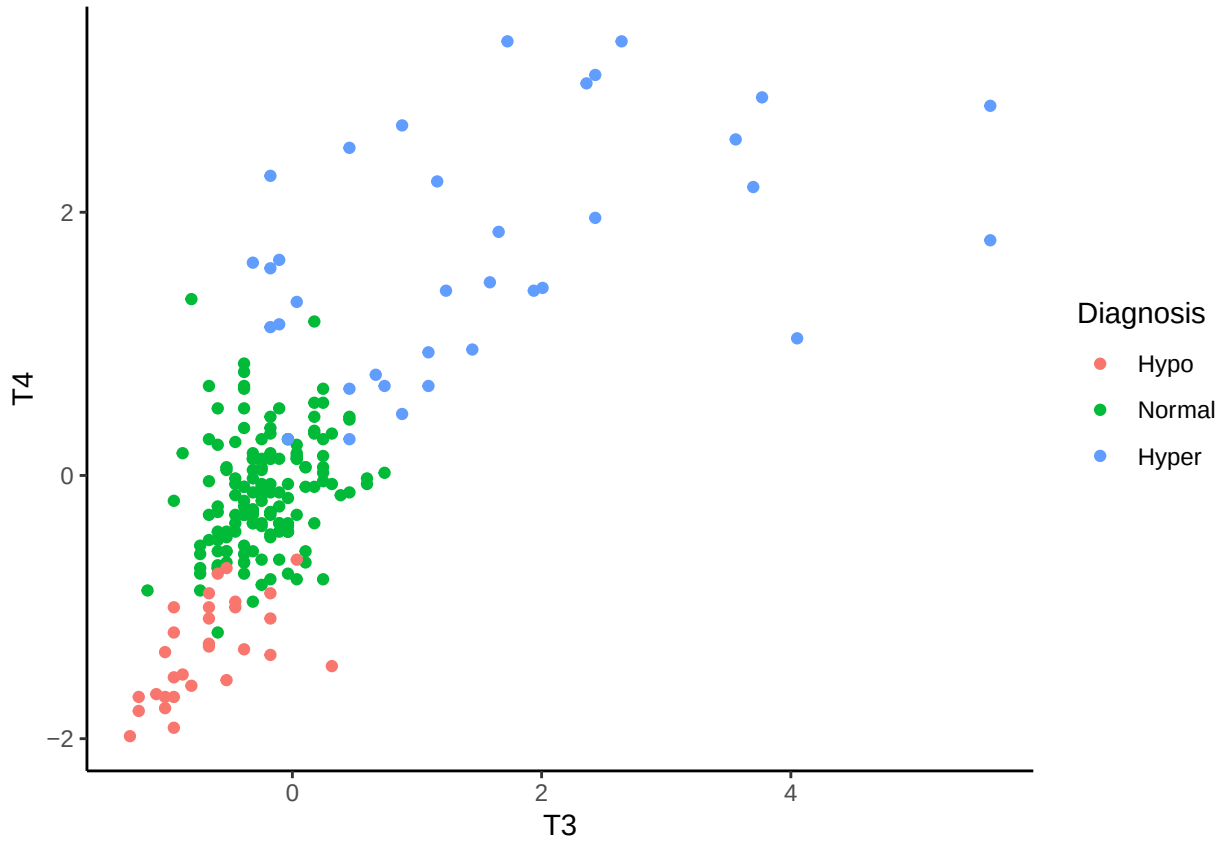
```
# visualize data
raw_data = mclust::thyroid
exposure = data.frame(scale(raw_data[,2:6])) # center and scale exposure data
exposure$Diagnosis = factor(raw_data$Diagnosis)

# 5d plot?? or just do multiple 2d plots ## should this data be log transformed
# before centering and scaling???
ggplot2::ggplot(data = exposure, mapping = aes(x = TSH, y = T3)) +
  ggplot2::geom_point(aes(color = Diagnosis)) +
  ggplot2::theme_classic()
```



```
ggplot2::ggplot(data = exposure, mapping = aes(x = TSH, y = T4)) +
  ggplot2::geom_point(aes(color = Diagnosis)) +
  ggplot2::theme_classic()
```




```
ggplot2::ggplot(data = exposure, mapping = aes(x = T3, y = T4)) +
  ggplot2::geom_point(aes(color = Diagnosis)) +
  ggplot2::theme_classic()
```




```
# put exposure data into format that model will accept
y = lapply(X = 1:nrow(exposure),
           FUN = function(x){
             matrix(exposure[x,], nrow = length(exposure[x,]))
           })
```

Model Fitting

```
# fit model

mod1 = MVN_CRP_sampler_DEV(
  S = 12000, seed = 516, y = y,
  alpha = 1, r = 10, g = 1, h = 50,
  sigma_hyperprior = FALSE, fix_r = FALSE,
  mu0 = matrix(rep(0, ncol(exposure)), ncol = 1),
  a = 1, b = 50,
  k_init = 1, diag_weights = FALSE,
  verbose = TRUE, split_merge = TRUE)


saveRDS(object = mod1, file = "../MCMC_Runs/thyroid_DEV_modfit.rds")
```

Model Summary

```
# summarize model fit
mod1 = readRDS("../MCMC_Runs/thyroid_DEV_modfit.rds")
mod1_sum = dpmm_summary(output = mod1,
                        print_phi_sum = TRUE,
                        print_k_sum = TRUE,
                        make_traceplot = TRUE,
                        burn_in = 2000, t_hold = 250,
                        num_dims = 5,
                        calc_perf = FALSE,
                        equal_var = FALSE)
```

```
##
##  Frequency of MCMC iterations finding K groups:
##    1    2    3    4    5    6    7    8
##  657 8827 2226  249   28   10    1    2
##
##  Percentage of MCMC iterations finding K groups:
##    1    2    3    4    5    6    7    8
##  5.5 73.6 18.6  2.1  0.2  0.1  0.0  0.0
##
##  *Note that above frequency summaries of MCMC iterations were made before burn-in or thresholds were
##           All inference on phi will be made after accounting for burn-in and thresholding.
##
##  K = 1  n_k = 541 after burn-in and thresholding
##        Mean Median Empirical SE  2.5% 97.5%
## mu_1_1    0   0.00            0 -0.08  0.09
## mu_1_2    0   0.00            0 -0.09  0.09
## mu_1_3    0   0.01            0 -0.10  0.10
## mu_1_4    0   0.00            0 -0.09  0.10
## mu_1_5    0  -0.01            0 -0.10  0.09
##           Mean Median Empirical SE 2.5% 97.5%
## sigma_1_1 1.09   1.08            0 0.99  1.18
##
##  K = 2  n_k = 7305 after burn-in and thresholding
##         Mean Median Empirical SE  2.5% 97.5%
## mu_1_1  0.02   0.02         0.04 -0.27  0.30
## mu_1_2 -0.08  -0.08         0.04 -0.38  0.20
## mu_1_3 -0.18  -0.18         0.04 -0.46  0.12
## mu_1_4 -0.23  -0.24         0.03 -0.52  0.07
## mu_1_5 -0.19  -0.20         0.03 -0.47  0.09
## mu_2_1 -0.11  -0.07         2.16 -1.31  0.88
## mu_2_2  0.27   0.31         1.19 -1.27  1.15
## mu_2_3  0.71   0.62         2.39 -0.58  1.96
## mu_2_4  1.02   0.82         2.00  0.30  3.40
## mu_2_5  0.86   0.66         1.68  0.17  2.90
##           Mean Median Empirical SE 2.5% 97.5%
## sigma_1_1 0.39   0.32         0.09 0.28  0.93
## sigma_2_1 4.84   3.68       699.99 2.68  9.35
##
##  K = 3  n_k = 1900 after burn-in and thresholding
##         Mean Median Empirical SE  2.5% 97.5%
## mu_1_1  0.03   0.02         0.04 -0.26  0.34
```

```
## mu_1_2 -0.09  -0.08         0.05 -0.40  0.23
## mu_1_3 -0.18  -0.19         0.06 -0.49  0.15
## mu_1_4 -0.24  -0.24         0.07 -0.54  0.07
## mu_1_5 -0.19  -0.20         0.06 -0.52  0.12
## mu_2_1 -1.05  -1.54        93.81 -5.85  4.18
## mu_2_2  1.99   1.94        66.27 -2.82  5.74
## mu_2_3  2.24   2.30        48.45 -2.34  7.37
## mu_2_4  0.23  -0.23        29.72 -1.95  8.02
## mu_2_5 -0.25  -0.40        23.10 -2.96  5.62
## mu_3_1  0.82   0.99         1.88 -1.49  2.11
## mu_3_2 -0.83  -1.00         2.22 -2.39  1.71
## mu_3_3 -0.34  -0.49         2.00 -1.66  2.17
## mu_3_4  1.97   1.93         2.46  0.27  5.47
## mu_3_5  1.79   1.74         2.93  0.10  4.71
##            Mean Median Empirical SE 2.5% 97.5%
## sigma_1_1  0.42   0.32         0.19 0.27  1.03
## sigma_2_1 31.24   2.13    360974.26 1.39 68.95
## sigma_3_1  4.81   2.90       116.86 2.06 18.92
##
##  Split/Merge MH Steps:
## # A tibble: 2 x 3
##   move_type Accept_Prob Count
##   <fct>           <dbl> <int>
## 1 MERGE           0.999   861
## 2 SPLIT           0.015  1539
##
##  Summary function runtime is 0.2587352 mins
```

```
mod1_sum$settings
```

```
## $S
## [1] 12000
##
## $alpha
## [1] 1
##
## $a
## [1] 1
##
## $b
## [1] 50
##
## $mu0
##      [,1]
## [1,]    0
## [2,]    0
## [3,]    0
## [4,]    0
## [5,]    0
##
## $k_init
## [1] 1
##
## $d
## [1] 1
```

```
##
## $f
## [1] 1
##
## $g
## [1] 1
##
## $h
## [1] 50
##
## $r
## [1] 8.460454
##
## $mod_type
## [1] "conjDEV"
##
## $split_merge
## [1] TRUE
##
## $sm_iter
## [1] 5
```

```
mod1_sum$fit_runtime
```

```
## Time difference of 197.3917 mins
```

```
mod1_sum$splitmerge_accept
```

```
## # A tibble: 2 x 3
##   move_type Accept_Prob Count
##   <fct>           <dbl> <int>
## 1 MERGE           0.999   861
## 2 SPLIT           0.015  1539
```

```
# save model summary
# saveRDS(object = mod1_sum, file = "../MCMC_Runs/thyroid_DEV_modsum.rds")
```